University of Sussex

Statistics of rare objects and the Integrated Sachs-Wolfe effect in ΛCDM N-body simulations

William A. Watson

Submitted for the degree of Doctor of Philosophy

University of Sussex

31st May 2013

Declaration

I hereby declare that this thesis has not been and will not be submitted in whole or in part to another University for the award of any other degree.

Parts of this thesis have been undertaken in collaboration with other researchers. Ilian Iliev performed the simulations that this work is based on. Jose Maria Diego produced Figures 4.5–4.7 and worked out the framework to calculate the lensing signal in the Jubilee simulation laid out in § 5.2.3. Stefan Gottlöber ran his void finder on the AHF halo data from the Jubilee simulation to produce the void catalogues in § 4.3.5. Airam Marcos-Caballero undertook the linear theory calculation for the ISW power spectra in Figure 5.8. Sesh Nadathur provided the filtered data upon which Figure 5.15 is plotted.

Signature:

William A. Watson

For my family

UNIVERSITY OF SUSSEX

William A. Watson, Doctor of Philosophy

Statistics of rare objects and the Integrated Sachs-Wolfe effect in ΛCDM N-body simulations

Summary

In this thesis we make predictions of extreme elements of large-scale structure (LSS) in the universe. We base our study on the concordance cosmological model, the Lambda Cold-Dark-Matter (ΛCDM) model, and in doing so we utilise a suite of very large N-body, dark-matter-only simulations.

To understand LSS throughout cosmic history, it is vital to quantify the evolution of the numbers of objects in the universe. To this end, we perform a numerical investigation into the abundance of dark matter haloes across an unprecedented combination of redshifts and masses. For the very young universe ($z > 6$), a fit is presented for the numbers of rare haloes that hosted the energetic objects that drove reionization. At lower redshifts we predict number counts of galaxy groups and clusters, the observation of which forms perhaps our current, best method of interpreting nature on large scales.

Our low redshift results are based on simulations with very large volumes, which allows us to probe rare objects in a ΛCDM universe, including massive clusters, voids and extreme-velocity mergers. These objects challenge our understanding of the universe by exhibiting the extremes of the ΛCDM model. With the possible exception of the Bullet Cluster, our simulation results are in line with current observations.

We study the late-time Integrated Sachs-Wolfe (ISW) effect using a $(6\ h^{-1}\mathrm{Gpc})^3$ volume simulation which contains enough particles (6000^3) to resolve luminous red galaxies. From these data we calculate the expected ISW-LSS cross-correlation signal in a ΛCDM universe. The signal is found to be strongest for LSS surveys that can probe redshift ranges of $z \sim 0.2$ to 0.8. The ISW effect promises to be an important measure of the evolution of dark energy, the overall understanding of which is perhaps the most important current goal in cosmology.

Acknowledgements

This thesis would not have been possible without the continual, steady, pragmatic support of my supervisor, Ilian Iliev. I owe him my deepest thanks for his excellent guidance and for his endless patience throughout my studentship. Special thanks must also be given to him for devising and running the simulations on which this work is based.

There are many others who have helped me in my research whom I shall now endeavour to thank. Jose Maria Diego and Gustavo Yepes first and foremost for their guidance and support; Alexander Knebe, Stefan Gottlöber, Paul Shapiro and Anson D'Aloisio with whom it has been a pleasure to collaborate; Steffen Knollman in particular for his part in solving the problem of how to reduce the vast data from our simulations into a more manageable form; Seb Oliver and Peter Thomas for helpful feedback and guidance; Leonidas Christodoulou, Leon Baruah, Donough Regan, Philip Rooney and Owain Young for many helpful discussions; Aurel Schneider for excellent feedback on much of the work in this thesis; Chris Short for his patient role as a help-desk for all things simulation-related and Shaun Hotchkiss, Sesh Nadathur and my collaborators at IFCA.

I would like to thank the support staff at the Texas Advanced Computing Center and the Jülich Supercomputing Centre for their help with the running and analysis of the simulations in this work.

Finally I would like to thank my fiancée Emina Demiri, my parents and my sister for their unwavering support, as well as the Demiri-Peirce-Dostal clan and the Hoskings for keeping me sane(ish) over the last few years. These are the 'family' to which I refer in the dedication of this work. Thank you all.

This research was funded by the South East Physics network (SEPnet[1]). Some of the results in this thesis have been derived using the HEALPix[2] package (Górski et al., 2005). Figures have been produced using HEALPix, IDL and Pyxplot[3].

[1] http://www.sepnet.ac.uk

[2] http://healpix.jpl.nasa.gov

[3] Which I heartily recommend.

Contents

List of Figures

To see a world in a grain of sand,
And a heaven in a wild flower,
Hold infinity in the palm of your hand,
And eternity in an hour.

– William Blake

Chapter 1

Introduction

And it seems to me that the ways by which men acquire knowledge of the celestial things are hardly less wonderful than the nature of these things themselves

– Johannes Kepler

1.1 Cosmology: a historical context

Today the field of cosmology is concerned with the universe on extremely large scales. The history of humanity's accumulation of knowledge regarding the nature of the physical universe can be viewed, broadly, as an expansion of our conception of the size of that universe. Modern cosmology is based on multiple observational strategies that attempt to deepen our view of the universe by enriching the wealth of data we possess that describes it. We look further, more broadly and more carefully at the universe with each passing decade of research.

An important area in cosmology, possibly the most important, is the investigation of the distributions of galaxies in the universe. These objects, observed throughout history, were a topic of debate for many years: did they lie within our galaxy or were they to be found further afield? As early as 1750 it was proposed by Thomas Wright that the Milky Way was a flat layer of stars and that the nebulae observed in the sky were individual galaxies in their own right. The nature of the nebulae was clarified in 1922 by Ernst Öpik, who calculated the distance to Andromeda (Opik, 1922) finding a value that was too distant to lie within the estimated bounds of our own galaxy. Further confirmation

of the enormous distances to the nebulae was given by Edwin Hubble in 1929 with his observations of Cepheid variable stars in Andromeda and other nearby galaxies (Hubble, 1929a,b). Although, interestingly, his measured value for the distance to Andromeda of 275 kpc was actually further off than Öpik's 450 kpc (the value we measure today is 778 ± 33 kpc (Karachentsev et al., 2004)). Given the newly-understood status of nebulae as 'Island Universes' (a term coined by Immanuel Kant), Hubble made the major leap forward of observing enough galaxies to ascertain the fact that the universe is expanding (Hubble, 1929a). This was observational confirmation of a proposed theoretical solution to Einstein's field equations in General Relativity, put forward by Georges Lemaître a few years before Hubble's discovery (Lemaître, 1927). As an empirical observation concerning the nature of the world in which we find ourselves it is quite possibly the most startling and important discovery humankind has made[1].

Hubble was able to plot a relationship between the distances from us to galaxies and their velocities. Except for a few galaxies that lay very close to the Milky Way, all were found to be moving away from us with recession velocities that are roughly proportional to their distances. This has since become known as the Hubble Law:

$$v = H_0 d, \tag{1.1}$$

where v is a galaxy's velocity, d is its distance and H_0 is the Hubble constant, which defines the slope of the velocity-distance relationship. The Hubble constant, which describes how quickly the present-day universe is expanding, constitutes one of the most important measures in cosmology, even today, almost a century after it was first measured.

Throughout the 20^{th} century much research in cosmology was focussed on mapping and measuring greater and greater numbers of galaxies. One of the main aims of these efforts was to more accurately measure H_0. Galaxies were typically found in groups and clusters. Gradually, a picture of the large-scale structure (LSS) of the universe was revealed. This picture showed that the universe, on the scale of tens to hundreds of Mpc, contains rich structure, with galaxies aggregating in walls, filaments and knots around huge, underdense void regions.

As early as the 1930s (Zwicky, 1933) studies of galaxy motions in clusters were suggesting that a great deal of matter in the universe must be invisible to astronomers and

[1]Hubble was not awarded a Nobel prize for his work, because at that time Astronomy was not considered to be an area of physics, rather a research field in its own right.

therefore must interact primarily only through gravity[2]. This matter was called *dark matter*. Today we model the evolution of dark matter in large N-body simulations and produce the same patterns of large-scale structure that we observe in the universe around us. This thesis is based on the outputs from a suite of such simulations.

A major observational breakthrough occurred in 1965 with the detection of the Cosmic Microwave Background (CMB) radiation (Penzias & Wilson, 1965). This radiation was produced around 380,000 years after the big bang due to the universe cooling enough to allow atoms to form without disturbance from ionising photons. The CMB, which has streamed through the universe since this point, passing peaks and troughs in the matter field along the way, is very homogenous. Since its original discovery this radiation has been carefully studied by a variety of satellite-, balloon- and ground-based telescopes. Small temperature anisotropies in the radiation, first revealed by the COBE mission (Smoot et al., 1992), reflect the density distribution in the very early universe. Figure 1.1 shows a Mollweide projection of these anisotropies from the recent Planck mission (Planck Collaboration, 2013a).

In 1998 it was discovered that the late-time universe was accelerating in its expansion (Riess et al., 1998; Perlmutter et al., 1999). This was a surprising result, found via datasets of type Ia supernovae that appeared dimmer than expected, and it pointed to a large component of the present-day universe existing in the form of a material with a negative pressure, termed *dark energy*.

1.2 Modern cosmology

Today we place bounds on how old the universe is to the third decimal point: 13.798 ± 0.037 billion years (Planck Collaboration, 2013a). Combined with a finite speed of light, this sets a precise measurement for the size of the universe that we can possibly observe around us. This measurement is one of the triumphs of modern cosmology. To put it in perspective, only 15 years ago we were still unsure of the values of the fundamental parameters of the universe. For example, the existence of dark energy was unknown until 1998. Nowadays the phrase 'precision cosmology' is widely-used in earnest, something that would have been unthinkable 20 years ago.

With the WMAP and Planck missions (Hinshaw et al., 2012; Planck Collaboration, 2013a), combined with multi-wavelength galaxy and cluster surveys (Condon et al., 1998;

[2]Another conclusion that can be reached from these observations is that the theory of gravity itself may require modification.

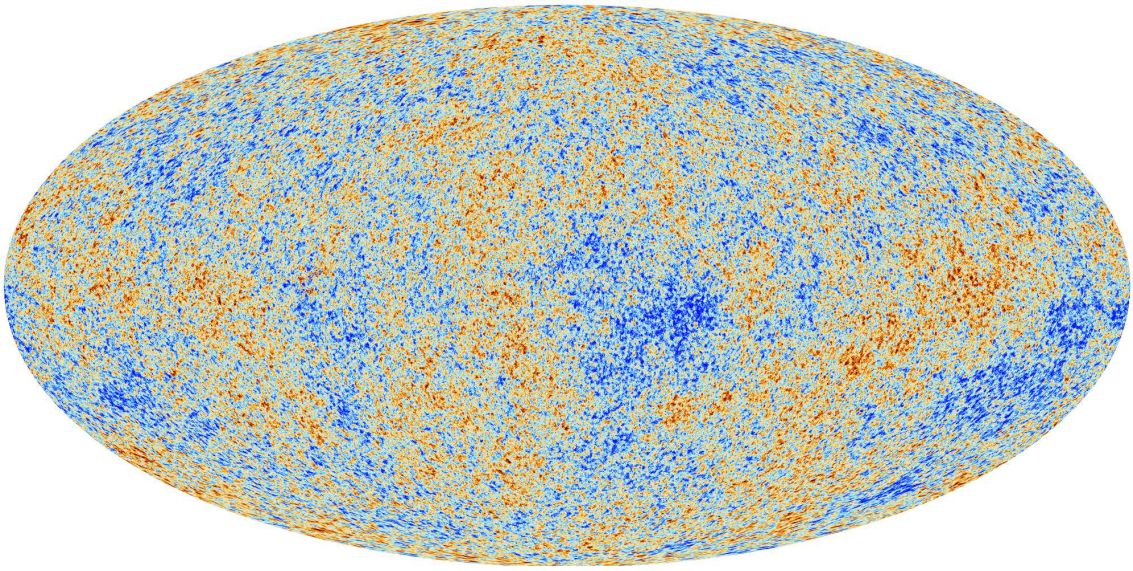


Figure 1.1: The Planck satellite CMB sky, showing anisotropies in the temperature of the cosmic microwave background radiation. Red patches show hot spots with a maximum deviation from the mean of around one part in a thousand. Blue regions show cold spots with a similar amplitude. *Credit: ESA Planck Collaboration.*

Böhringer et al., 2004; Burenin et al., 2007; Healey et al., 2007; Eales et al., 2010; Murphy et al., 2010; Heymans et al., 2012; Anderson et al., 2012) and data from supernovae (Guy et al., 2007; Holtzman et al., 2008; Kessler et al., 2009; Guy et al., 2010; Conley et al., 2011; Suzuki et al., 2012), the cosmological model that best describes our universe – the ΛCDM model in the title of this thesis – has been pinned down with unprecedented accuracy. This is a model that describes the universe as constituted of mainly dark energy (Λ) and cold dark matter (CDM). Structure arises from primordial perturbations in the density field of the universe that were seeded during a time of rapid expansion in its very early history known as the inflationary epoch. Remarkably, the ΛCDM model is able, with just 6 parameters[3], to match data from multiple complementary cosmological probes. It is described as the 'concordance' model for this very reason. Given the success of this model in describing the universe, the inevitable question becomes 'where next for cosmology'?

In fact, it would be a dangerous attitude indeed to assume that we have a thorough understanding of the universe we see around us today just because various cosmological parameters are now quotable to a few decimal places. It can be said of ancient astronomers

[3]They are: the Hubble constant; the present-day values of the amount of dark energy and dark matter; the amplitude of density fluctuations; the tilt of the primordial power spectrum and finally the optical depth due to reionization. We will cover some of these parameters in detail in Chapter 2.

that they were tremendously ignorant of the nature and scale of the universe. Our situation today is not too dissimilar: we have no precise understanding of what the majority of the universe *is* as measured by density. Dark matter and dark energy are calculated to account for a total of 95% of the constituent material in the universe. We can see their effect on the universe and thereby characterise, broadly, their properties but we have no grasp of what they are at the fundamental level. This, perhaps, makes our situation more tractable than the ancients: we know what we do not know and we know we must look into the mystery of the dark sector to understand the universe more completely.

To that end, cosmology today is keenly focussed on the challenge of understanding dark energy and dark matter. Direct searches for dark matter are taking place in many detector experiments (see Bergström (2012) for a review, and references therein for details of the experiments) with the hope of finding dark matter particle candidates via their interaction through the weak nuclear force. A host of observational surveys are being undertaken to better investigate the nature of dark energy (Dawson et al., 2013; Drinkwater et al., 2010; Schlegel et al., 2009; Magnier et al., 2013; Mohr et al., 2012; Benítez et al., 2009; LSST Dark Energy Science Collaboration, 2012; Amiaux et al., 2012).

The first major question that researchers are attempting to answer regarding dark energy is what its equation of state is, i.e. what is its density-pressure relationship. Other important questions currently being addressed in studies of dark energy include whether it is spatially varying and whether it has been evolving over time. The question of what, physically, dark energy is has been only modestly addressed so far. An attempt to quantify it by identifying it with vacuum energy, as calculated by Quantum Field Theory, has resulted in a predicted value of dark energy that is incorrect by ~ 120 *orders of magnitude*. This has been widely referred to as the worst theoretical prediction ever made. It is unsurprising, therefore, that the direct detection of dark energy in a laboratory experiment has received little attention, not least due to its incredibly low density ($\sim 6\times10^{-27}$ kgm^{-3}). The investigation of dark energy in the near future therefore looks set to take place solely in the arena of astronomy.

1.2.1 Abundances of objects

The abundance of objects such as galaxies, groups and clusters of galaxies – and at earlier times quasars, proto-galaxies and even the very first stars – is an important measure in cosmology. We associate most of these objects spatially with peaks in the density field of the universe. As the majority of matter exists in the form of dark matter we describe

these density peaks as *dark matter haloes*. The halo mass function, which describes how counts of these dark matter haloes vary with mass, plays an important role in cosmology due to its sensitivity to a number of important parameters, including the matter density of the universe, the Hubble parameter, the spectral index of the primordial power spectrum and the dark energy equation of state (e.g. Holder et al., 2001; Haiman et al., 2001; Weller et al., 2002). An important prediction of the ΛCDM model is therefore the expected number counts of dark matter haloes across cosmic time. The mass function is currently difficult to pin down with much precision observationally (for recent results see Rines et al. (2008); Vikhlinin et al. (2009); Rozo et al. (2010) and for an overview of the observational challenges see Eke et al. (2006)). However, we can investigate it using N-body simulations to model the growth of structure on large scales.

The mass function is employed widely in cosmology. For the late-time universe it is used in statistical analyses of cluster surveys to constrain cosmological parameters (e.g. Smith & Marian (2011)). These large objects probe the high mass end of the function, which is more sensitive to cosmological parameters than the lower mass end (for a recent review paper on the formation of galaxy clusters, including the role of the mass function, see Kravtsov & Borgani (2012)). The mass function is a key component in studies employing correlations between galaxies, as the halo-halo term in the correlation function directly depends on the mass function (see for example Cooray & Sheth (2002) and references therein). For earlier times in the history of the universe the mass function is applied in modelling the process of reionization, which was driven by sources residing in dark matter haloes including population-III stars, early galaxies and accreting black holes. This was a crucial epoch in the universe's history of which we hope to make precision observations in the coming years using the Square Kilometre Array and its pathfinder projects (Rawlings, 2011; Fan, 2012). The effect of reionization on the photons of the CMB radiation, specifically the optical depth of the ionised universe, is one of the 6 parameters we mentioned above that make up the ΛCDM model's description of the universe. It is very important, therefore, to understand what the number-counts of objects were during this era, since they were responsible for driving the reionization process.

Any significant, observed deviation from the mass function predicted by the ΛCDM cosmological model would create tension in our current understanding of structure formation in the universe. For example, we are able to put bounds on how massive the largest observable clusters in the visible universe should be from the mass function (see Harrison & Coles, 2012, and references therein), the discovery of larger clusters would put tension

on the ΛCDM model.

1.2.2 The Integrated Sachs-Wolfe effect and gravitational lensing

As well as directly using non-linear structures and their number statistics as a function of time (or even simply their mere existence) to gain deeper insight into cosmological models, we can also study their imprint on observations of the CMB. The Planck satellite (Planck Collaboration, 2013a) has observed the sky in the microwave band with frequencies ranging from 30 GHz and up to 857 GHz. The main goal of Planck is to measure the CMB with unprecedented quality over the entire sky. One of the most interesting applications of the Planck CMB data is the detection (Planck Collaboration, 2013b) of the Integrated Sachs-Wolfe effect (ISW), which is the result of the balance between the gain and loss of energy of the CMB photons when they fall into and subsequently leave the gravitational potentials, i.e. the potentials generated by structures in the universe. In a universe with no dark energy, this energy balance cancels out but in a universe dominated by dark energy, this energy balance results in a net positive energy gain for the CMB photons when they cross overdense regions and a negative one when they cross underdense regions. The effect is at its most pronounced at large angular scales and late times, where and when the dark energy component of the cosmological fluid has its greatest influence on the shape of the potentials. A detection of the ISW is thus a measurement of dark energy. Since the ISW is frequency independent (like the CMB itself), its detection is possible only through a cross-correlation analysis with other datasets that trace the potential (such as the aforementioned observational surveys). This necessitates the application of large enough observational volumes to properly account for all relevant scales and the impact of cosmic variance.

Detailed studies using simulations of the expected cross-correlation signal between the large-scale structure of the universe and the anisotropies in the CMB radiation induced by the ISW effect are very difficult to undertake. This is because the simulations need to have very large volumes to capture the large scale nature of the ISW effect whilst also being able to resolve structures such as galaxy-hosting haloes accurately. The balance between these two requirements results in a need for simulations to have an extremely large number of particles (hundreds of billions), which makes them taxing to implement.

The ISW effect arises from the dark energy driven time variability of the gravitational potential. This effect occurs in the radial direction of the potential as we view it. There are also variations that occur in the tangential direction. These show up as the gravita-

tional lensing effect, which preserves the temperature of photons but distorts their paths. These distortions concentrate on small scales (of the order of a few arcminutes) and hence complement the large-scale ISW. The effect is very sensitive to the mass distribution in the universe and as such provides a method of probing cosmological parameters (Mandelbaum et al., 2013). Gravitational lensing will be a source of confusion noise in future CMB polarisation missions (like the proposed COrE mission (The COrE Collaboration et al., 2011)). Large simulations are needed to properly account for this source of systematic error and to study ways of reducing its impact. Among these projects, future space missions such as Euclid (Amiaux et al., 2012) will need to rely on realistic simulations that include not only the lensing effect due to large scale structure but also the associated catalogues that trace that matter. Simulations will be needed to validate the methods employed in these missions.

1.3 This thesis

This thesis quantifies various predictions of the ΛCDM model using numerical simulations. The work is based on a suite of very large (with $N_p = 3072^3 - 6000^3$ particles) cosmological N-body simulations, all of which were run by my supervisor, Ilian Iliev, based on the ΛCDM cosmological model.

Thus far there have been many studies on the mass function but few have covered a broad swathe of cosmological history and most have neglected the function for early times in the universe. We address this issue and provide predictions for the abundance of dark matter haloes across an unprecedented range of cosmic time.

We use one of the largest dark matter simulations performed to-date – the Jubilee simulation – to calculate the large-scale ISW effect from the structures in the simulation volume. We also calculate the weak lensing effect induced by the same structures and produce all-sky maps of both. We then investigate the ISW-LSS cross-correlation signal, using mock galaxy catalogues and the ISW maps. From the large volume of the Jubilee simulation we also investigate some of the statistics of rare objects in the simulation, including the most massive galaxy clusters, and provide a measure of how cosmic variance affects cluster counts.

In Chapter 2 we present an overview of the ΛCDM model, including briefly describing its theoretical underpinnings. We also cover background on N-body simulations in general and provide detail on the particulars of the simulations and algorithms used in this work.

In Chapter 3 we investigate how the dark matter halo mass function evolves with

time. We also look in some detail at how different definitions of halo mass affect the mass function results.

In Chapter 4 We present results from the Jubilee simulation, a large N-body, dark matter-only cosmological simulation with a volume of $V = (6\ h^{-1}\mathrm{Gpc})^3$, containing 6000^3 particles. We look at the statistics of the rarest clusters, merging systems and voids in the simulation and derive the effect of cosmic variance on cluster counts.

In Chapter 5 we study the expected large-scale ΛCDM ISW effect in the Jubilee simulation. We also calculate the gravitational lensing effect, which is created by the same potential as the ISW. We calculate the cross-correlation between the ISW effect and large-scale structure by constructing mock catalogues of Luminous Red Galaxies.

We give concluding remarks in Chapter 6.

Chapter 2

Theory

Philosophy is written in this immense book that stands ever open before our eyes (I speak of the universe), but it cannot be read if one does not first learn the language and recognise the characters in which it is written.

– Galileo Galilei

This chapter provides a brief overview of the theoretical underpinnings of the work in this thesis, beginning with a summary of the main results from General Relativity that pertain to cosmology. It also deals with the particulars of the ΛCDM concordance model, including an overview of how large-scale structure grows in such a model using linear perturbation theory. We then give a brief introduction into the concept of a power spectrum before finally outlining N-body simulations in general and the specifics of the simulations used in this work.

2.1 Einstein's theory of General Relativity

In the quotation at the head of this chapter Galileo alludes to mathematics as the language in which the book of nature was written and goes on to say: *It is written in the language of mathematics and its characters are triangles, circles and other geometrical figures, without which it is humanly impossible to understand a single word of it; without these, one is wandering around in a dark labyrinth.* For the purposes of this chapter, the quotation was abridged to imply that the language in question, when open to interpretation, could be a specific area of the mathematical sciences: that of Einstein's theory of General Relativity, for this is the language in which the universe on large scales is written. Galileo's comment is peculiarly pertinent to modern cosmology, as much of our understanding of the universe on the very largest scales is geometrical in nature. General Relativity describes reality

as a 4-dimensional space-time. The evolution of our universe then takes the form of a geometrical picture with a simple parameterisation in time that captures the expansion of its spacial part. Before we move onto specifics we shall cover some aspects of the theory of General Relativity in which we loosely follow Hobson et al. (2006).

The theory rests on the description of arbitrarily curved space-time manifolds via their metrics, such that a line element in an any given space-time takes the form:

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu, \tag{2.1}$$

where $g_{\mu\nu}$ is the metric of the space-time and we have assumed Einstein's summation convention, where repeated indices are implicitly summed-over. As the indices refer to space-time coordinates they run from 0-3 where the 0^{th} index refers to the time coordinate. In the following we define the signature of the metric to be (+, –, –, –).

The key result of Einstein's theory of General Relativity is the Einstein equation:

$$R_{\mu\nu} = -\frac{8\pi \mathrm{G}}{c^4}(T_{\mu\nu} - \frac{1}{2} T g_{\mu\nu}) + \Lambda g_{\mu\nu}, \tag{2.2}$$

where $R_{\mu\nu}$ is the Ricci tensor, $T_{\mu\nu}$ the energy-momentum tensor and $T = T^\mu_\mu$. Λ is the *cosmological constant* term. This tensor equation describes a system of 10 independent field equations. Solutions to these equations typically require some simplifying symmetries of the physical space-time in question to be used in order to reduce the complexity of the theory. For the case of the field equations of the universe we now turn to the Friedmann-Lemaître-Robertson-Walker (FLRW) solution.

2.1.1 The Friedmann-Lemaître-Robertson-Walker (FLRW) solution to the field equations

In order to simplify equation 2.2 to apply General Relativity to the universe as a whole, we need to make two assumptions about the nature of the universe. These are wrapped

into the *cosmological principle*, which states that

At any time, the universe looks the same from all positions in space, and all directions in space at any point are equivalent.

This principle implies that the universe is both *homogenous* (the same in all positions) and *isotropic* (the same in all directions). These two assumed properties then allow a simplified treatment of the field equations. To begin with we can reduce the metric of this type of space-time to

$$ds^2 = c^2\ dt^2 - a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\ \theta d\phi^2)\right], \tag{2.3}$$

where the evolution in time of the spatial part of the metric is now wholly contained in the *scale factor*, $a(t)$; k is the curvature of space, and r, θ and ϕ are the standard spherical polar coordinates, although, importantly, it should be noted that they are *comoving* coordinates, by which we mean that an observer with no peculiar velocity in the space-time has a fixed position in (r, θ, ϕ). Furthermore, *all* points in space in this metric are equivalent. There are no distinguished observers, therefore the origin of the coordinate system in this metric can be chosen to be in an arbitrary location.

We now have a metric to plug into the field equations (equation 2.2). From the metric it is straightforward (but rather tedious) to calculate the Ricci tensor, via the affine connection, and therefore we are left with the question of how to devise a suitable energy-momentum tensor for the universe. A typical approach is to take as simple a fluid as possible[1]. This fluid is one that is devoid of shear-viscous, bulk-viscous and heat-conductive properties, i.e. a *perfect fluid*. The energy-momentum tensor in this case is then

$$T^{\mu\nu} = \left(\rho + \frac{p}{c^2}\ u^\mu u^\nu\right) - pg^{\mu\nu}, \tag{2.4}$$

where the density, ρ, and pressure, p, can vary only in time so as to preserve the condition of homogeneity. It is then possible to calculate the Einstein equations for an homogenous, isotropic universe. These equations are called the cosmological field equations or the

[1] Although not to the extent of assuming there to be no fluid at all, as per the Milne model.

Friedmann-Lemaître equations:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho}{3} + \frac{\Lambda c^2}{3} - \frac{kc^2}{a^2}, \tag{2.5}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3}. \tag{2.6}$$

Equation 2.5 is of great importance because it governs the rate of expansion of the universe. The Hubble parameter is defined as

$$H(t) = \frac{\dot{a}}{a}. \tag{2.7}$$

Note that the Hubble constant, as expressed in equation 1.1, is the present day value of the Hubble parameter, $H_0 = H(t = 0)$.

From equation 2.5 we see that $H(t)$, which is the rate of expansion or contraction of the universe, is governed by the interplay of three terms. The first, a term driven by the density field of the universe, the second, a term driven by the cosmological constant, Λ, and the third driven by the spatial curvature, k.

By taking the time derivative of equation 2.5 we can eliminate $\ddot{a}$ from equations 2.5 and 2.6 and derive the cosmological equation for energy conservation:

$$\dot{\rho} + \left(\rho + \frac{p}{c^2}\right)\frac{3\dot{a}}{a} = 0. \tag{2.8}$$

The modern view of Λ is that it represents the energy density of the vacuum and as such it can also be described as a perfect fluid. To this end we consider the contents of the universe to consist of multiple components:

$$\rho(t) = \rho_m(t) + \rho_r(t) + \rho_\Lambda, \tag{2.9}$$

where ρ_m is the density of matter, including Cold Dark Matter (CDM) and baryons; ρ_r, the density of radiation, which we take to be the sum of the density of both photons and neutrinos; and ρ_Λ, the density of the vacuum energy (which here we assume is constant

in time). Each of the components can be modelled as perfect fluids in their own right and have equations of state of the form:

$$p_i = \omega_i \rho_i c^2, \tag{2.10}$$

where the subscript refers to the component under consideration and ω is the equation of state parameter for each fluid. For a pressureless fluid $\omega_m = 0$, which is the case for CDM. In the following we make the assumption that all[2] matter has this equation of state, including baryons. For radiation $\omega_r = \frac{1}{3}$ and for the vacuum $\omega_\Lambda = -1$. Any deviation from $\omega = -1$ would imply a deviation from a pure cosmological constant-type fluid. To represent this situation the vacuum energy fluid is often referred to as *dark energy*. Dark energy may have a value for ω that is not -1. It may also vary either in space or with time or both. In this thesis the term 'dark energy' is used to refer to the general case of a cosmological fluid with an equation of state parameter near $\omega \sim -1$ that may or may not vary. The term 'cosmological constant' is used to refer to the precise case of $\omega = -1$. The symbol Λ will be used to represent both cases and it is hoped that from the context it is clear which term is meant. Combining equations 2.8 and 2.10 gives

$$\rho_i(t) \propto a(t)^{-3(1+\omega_i)}. \tag{2.11}$$

This equation yields the evolution of the densities for the components of the universe:

$$\rho_m(t) = \rho_{m,0} \left(\frac{a(t)}{a_0} \right)^{-3}, \tag{2.12}$$

$$\rho_r(t) = \rho_{r,0} \left(\frac{a(t)}{a_0} \right)^{-4}, \tag{2.13}$$

$$\rho_\Lambda(t) = \rho_{\Lambda,0}, \tag{2.14}$$

[2]This ignores any contribution to the matter density from relativistic particles, where their thermal energy is comparable to their rest mass energy. Such is the case for warm dark matter, for example.

where the subscript '0' refers to present day values. For the case of the vacuum energy, we see from equations 2.5 and 2.14 that to wrap the Λ term into the density we require that

$$\rho_\Lambda(t) = \rho_{\Lambda,0} = \frac{\Lambda c^2}{8\pi \mathrm{G}}. \tag{2.15}$$

Using these densities we can then define the following parameters for the FLRW model solution:

$$\Omega_m(t) \equiv \frac{8\pi \mathrm{G}}{3H^2(t)}\rho_m(t), \tag{2.16}$$

$$\Omega_r(t) \equiv \frac{8\pi \mathrm{G}}{3H^2(t)}\rho_r(t), \tag{2.17}$$

$$\Omega_\Lambda(t) \equiv \frac{8\pi \mathrm{G}}{3H^2(t)}\rho_\Lambda(t) = \frac{\Lambda c^2}{3H^2(t)}. \tag{2.18}$$

These parameters allow us to re-write equation 2.5 in the form:

$$1 = \Omega_m + \Omega_r + \Omega_\Lambda - \frac{kc^2}{H^2a^2}. \tag{2.19}$$

Defining a curvature density parameter as

$$\Omega_k(t) \equiv -\frac{kc^2}{H^2a^2}, \tag{2.20}$$

we have the very simple relation:

$$1 = \Omega_m + \Omega_r + \Omega_\Lambda + \Omega_k. \tag{2.21}$$

This equation dictates how the density balance of the contents of the universe evolves over time. In considering FRLW models it is useful to define the *total density parameter*:

$$\Omega \equiv \Omega_m + \Omega_r + \Omega_\Lambda = 1 - \Omega_k. \tag{2.22}$$

The motivation behind this is to describe simply the three possible cases for the curvature of an FLRW universe. For $\Omega < 1$ the universe has negative spatial curvature and we call the universe 'open'. For $\Omega = 1$ the universe has no curvature and we call the universe 'flat'. Finally, for the case where the universe has $\Omega > 1$ it has positive curvature and we call it 'closed'. We can then define the *critical density*, which is the density of the universe required to make it geometrically flat:

$$\rho_{crit} \equiv \frac{3H^2}{8\pi \mathrm{G}}. \tag{2.23}$$

Current observations suggest that our universe has a density that is very close to the critical value (see § 2.1.3).

2.1.2 Light propagation and distance measures in FLRW models

A topic of major importance in cosmology is that of how light rays travel across a space-time that is evolving. Observational astronomy and cosmology consists almost entirely of observing light that has propagated through the universe to us. To understand what we see in the sky we first need to understand how this light has been affected on its journey to us. In doing so it is then possible to accurately measure distances in the universe.

2.1.2.1 Cosmological redshift

Light rays travelling through the universe towards us experience an effect called *cosmological redshift*. This is the increase of the wavelengths of photons as they travel through a space that is expanding. There are many ways to derive the formula for cosmological redshift. Here we consider two galaxies, separated in space and occupying fixed comoving coordinates. We are free to choose any origin we wish for the coordinate system of the FLRW model so we define the origin to be at the location of one of the galaxies.

We now consider the emission of a photon at time t_E from the second galaxy directly towards the galaxy at the origin of the coordinate system. This photon travels radially, such that $d\theta = d\phi = 0$. Also for photons we have $ds = 0$, this results in equation 2.3 reducing to

$$c\ dt = a(t)\frac{dr}{\sqrt{1 - kr^2}}. \tag{2.24}$$

Now we can integrate for a photon's path from emission to reception:

$$\int_{t_E}^{t_R} \frac{c\ dt}{a(t)} = \int_{R_E}^{0} \frac{dr}{\sqrt{1-kr^2}}. \tag{2.25}$$

A second photon is now sent out from the emitting galaxy after a time δt_E has passed, such that it is emitted at $t_E + \delta t_E$ and received at time $t_R + \delta t_R$. We then have

$$\int_{t_E+\delta t_E}^{t_R+\delta t_R} \frac{c\ dt}{a(t)} = \int_{R_E}^{0} \frac{dr}{\sqrt{1-kr^2}} = \int_{t_E}^{t_R} \frac{c\ dt}{a(t)}, \tag{2.26}$$

which therefore implies

$$\int_{t_R}^{t_R+\delta t_R} \frac{c\ dt}{a(t)} = \int_{t_E}^{t_E+\delta t_E} \frac{c\ dt}{a(t)}. \tag{2.27}$$

Now, if δt_E and δt_R are small, then $a(t)$ can be assumed to be unchanged on both sides of equation 2.27. We consider the case where the photons are the pulses of successive wave crests of a light wave, making the time between their successive emissions and the time between their successive receptions very small compared to the evolution of the scale factor[3]. We thus have

$$\frac{\delta t_R}{a(t_R)} = \frac{\delta t_E}{a(t_E)}. \tag{2.28}$$

For successive wave crests we then define the ratio of the frequencies of the emitted and received photons to be the redshift:

$$1 + z \equiv \frac{\nu_E}{\nu_R} = \frac{\delta t_R}{\delta t_E} = \frac{a(t_R)}{a(t_E)}. \tag{2.29}$$

We usually define the scale factor at the present day to be 1. Therefore, for us, receiving light-rays on the Earth, there is a simple relation linking the cosmological redshift of objects and the scale factor:

$$a(t) = \frac{1}{1+z}. \tag{2.30}$$

[3]For the case of the very early inflationary universe this may not be the case, nor, potentially, in a late-time universe being dominated by certain types of dark energy.

That light is redshifted in this manner is highly fortunate as it allows us to define a distance to objects via their redshifts. We now consider some other distance-measures in FLRW universes.

2.1.2.2 Distance measures

The measurement of distances over cosmological scales is complicated by the expansion of the universe, its underlying geometry, and the finite speed of light (Hogg, 1999). In equation 1.1 we referred to a distance to a galaxy, d, without making a statement as to what we meant by that distance. In that particular case the distance referred to was the *proper distance* between a galaxy and the Earth. This distance is the measurement one would make if it were possible to instantaneously lay down a ruler between us and the galaxy to which we wish to know the distance.

We have already introduced the concept of comoving observers. These are observers whose spatial coordinates remain unchanged over time, i.e. observers who move with the Hubble expansion of the universe. We can define the *comoving distance* between these observers to be

$$d_C(z) = \int_0^z \frac{c\ dz'}{H(z')} = d_H \int_0^z \frac{dz'}{E(z')}, \tag{2.31}$$

where $d_H = c/H_0$ is the Hubble distance and $E(z)$ is the dimensionless Hubble parameter defined as

$$E(z) = \sqrt{\Omega_r(1+z)^4 + \Omega_m(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\Lambda}. \tag{2.32}$$

The comoving distance is unchanged between points that have no peculiar velocity with respect to the background Hubble expansion. The relationship between the proper distance and the comoving distance is simply

$$d = a(t)d_C. \tag{2.33}$$

These distance measures therefore coincide for the very near, low redshift universe and begin to diverge at early times (higher redshifts) as the evolution of the scale factor

becomes important. In cosmological simulations the comoving distance plays a central role as particle positions are typically described using a comoving coordinate system (§ 2.4.3).

The distance between two objects the same distance away from us but separated by some angle on the sky, $\delta\theta$, is $d_M \delta\theta$ where d_M is the *transverse comoving distance*. This distance is sensitive to the curvature of the FLRW model in question and is defined as (Hogg, 1999):

$$d_M = \begin{cases} d_H \ \Omega_k^{-\frac{1}{2}} \ \sinh\left[\Omega_k^{\frac{1}{2}} \ d_C/d_H\right] & \text{for } \Omega_k > 0 \\ d_C & \text{for } \Omega_k = 0 \\ d_H \ |\Omega_k|^{-\frac{1}{2}} \ \sin\left[|\Omega_k|^{\frac{1}{2}} \ d_C/d_H\right] & \text{for } \Omega_k < 0 \end{cases} . \tag{2.34}$$

Unfortunately, except in the very nearby universe, neither the comoving or proper distances are physically measurable. This is because as we look at objects that are very distant from us we are looking back in time to when the proper distances were smaller. As a result, we use more pragmatic distance measures observationally (and often just talk of an object's redshift to give a measure of how far away it is). The first of these is the *luminosity distance*, $d_{\rm L}$. This is the distance measured using the relation for the received flux, F, of an object of known luminosity, L:

$$d_{\rm L} = \left(\frac{L}{4\pi F}\right) = (1+z) d_M(z). \tag{2.35}$$

Note that it is necessary to have prior knowledge of the luminosity of a measured object in order to use this distance measure. Luminosity distance is used in particular for *standard candles*, objects in the sky that have known luminosities. These include the Cepheid variables employed by Hubble and type Ia supernovae, which were used in the first detections of dark energy (Riess et al., 1998; Perlmutter et al., 1999). Combining datasets such as these with other methods of ascertaining the distances to objects (for example using parallax to find the distances to nearby stars) allows us to construct the astronomical *distance ladder*, which is widely employed in cosmology, not least in trying to accurately measure the Hubble constant (Freedman et al., 2001).

Finally, we define the *angular diameter distance*, d_A. This distance measure supposes that there exists a standard ruler of known length, ℓ, whose angular diameter we can measure on the sky. Its angular diameter distance is then

$$d_A = \frac{\ell}{\Delta\theta} = \frac{d_M(z)}{1+z}. \tag{2.36}$$

Due to the $1/(1+z)$ dependence of this distance measure it has the property that it grows to a maximum value (near $z \sim 1$) and then diminishes past this point. Very distant objects can therefore subtend the same angular size on the sky as nearby objects.

An important point relating to these distance measures is that they are *cosmology dependent*. For different cosmological histories of the scale factor, $a(t)$, objects will have different values for the distance measures. Also of note is the fact that at low redshifts all the measurements, including the proper distance, converge to the same value.

2.1.3 The concordance ΛCDM FLRW cosmological model

It is assumed that the cosmological principle holds in our universe. This at first may seem counter-intuitive as there are obviously gross variations in the distribution of matter in the universe as we look out into the night sky. Isotropy is clearly not valid. Furthermore, we have no way of knowing what the universe looks like from other vantages. However, as we have conducted larger and larger surveys of the distribution of galaxies in the universe we have observed that the universe begins to become isotropic and homogeneous on large scales. These scales typically are of the order of hundreds of Mpc (Yadav et al., 2005; Sarkar et al., 2009; Sylos Labini & Baryshev, 2010).

Using observations of the CMB, LSS and the supernovae, we are able to pin down the parameters of the FLRW model that describes our universe. These parameters include: the age of the universe; the present day Hubble constant, H_0; the present day density parameters, Ω_i, including a density parameter for baryons, Ω_b; and the equation of state parameter for dark energy, ω_Λ. We show in Table 2.1 these values from the recent Planck mission Planck Collaboration (2013c). The dataset shown is the 'Planck+WP+highL+BAO' result, which combines observations of the CMB by the Planck and WMAP satellites with information from LSS surveys of how the Baryonic Acoustic Oscillations (BAOs), visible in the distribution of the scales of the anisotropies in the CMB (see § 2.2.1), show up at later times in the distribution of galaxies in the universe. This model, due to the domination of the cosmological constant, Λ, and CDM, has become known as the ΛCDM model. It is often referred to as the 'concordance' model because it can fit data across various complementary datasets, such as the CMB, LSS and supernovae data.

The first striking aspect of the results in Table 2.1 is their accuracy. Secondly, the

Parameter	Value
H_0	67.80 ± 0.77 kms^{-1}Mpc^{-1}
$100\Omega_k$	$-0.05^{+0.65}_{-0.66}$
Ω_Λ	0.692 ± 0.010
Ω_{CDM}	0.2584 ± 0.0037
Ω_b	0.04821 ± 0.00052
Age [Gyr]	13.798 ± 0.037
ω_Λ	$-1.13^{+0.24}_{-0.25}$

Table 2.1: ΛCDM parameters from the Planck Mission Planck Collaboration (2013c). The parameters are from the 'Planck+WP+highL+BAO' results, representing the best fit model for a combination of results from Plank, WMAP and LSS surveys.

curvature parameter is extremely close to zero, indicating that our universe is very close to being geometrically flat. The present day values of the density parameters illustrate the dominant Λ and CDM components. Currently, the best fit value for the dark energy equation of state parameter admits a solution of $\omega_\Lambda = -1$ with no tension. A cosmological constant is therefore a valid explanation for the late-time acceleration of the universe[4]. Given the known evolution of the different fluid components of the universe (equations 2.12–2.14) we can plot, for the ΛCDM model, what the expected evolution of the universe will look like, with regards to its constituents. We show this plot in Figure 2.1. For small values of the scale factor, which correspond to the high-redshift, early universe, radiation is the dominant component. Matter and radiation equality occurs at a redshift of $z_{eq} = 3391 \pm 60$ (based on the 'Planck + WMAP' value of Planck Collaboration (2013c)). After this point until very recently matter is the dominant component of the universe. This is an important result as it makes N-body simulations, which model the interactions of particles of matter via gravity, viable. At late times we see the rise of a dark energy component, calculated in Figure 2.1 as a cosmological constant. Dark energy models can show variations in this curve, depending on the time evolution of the model in question.

[4]Interestingly the best fit value for ω_Λ is slightly less than -1. If this is a true result (and there is no way of knowing with the accuracy level of the data), then the dark energy is described as *phantom energy* and the universe may well be headed for what has been termed a 'big rip' (Caldwell et al., 2003).

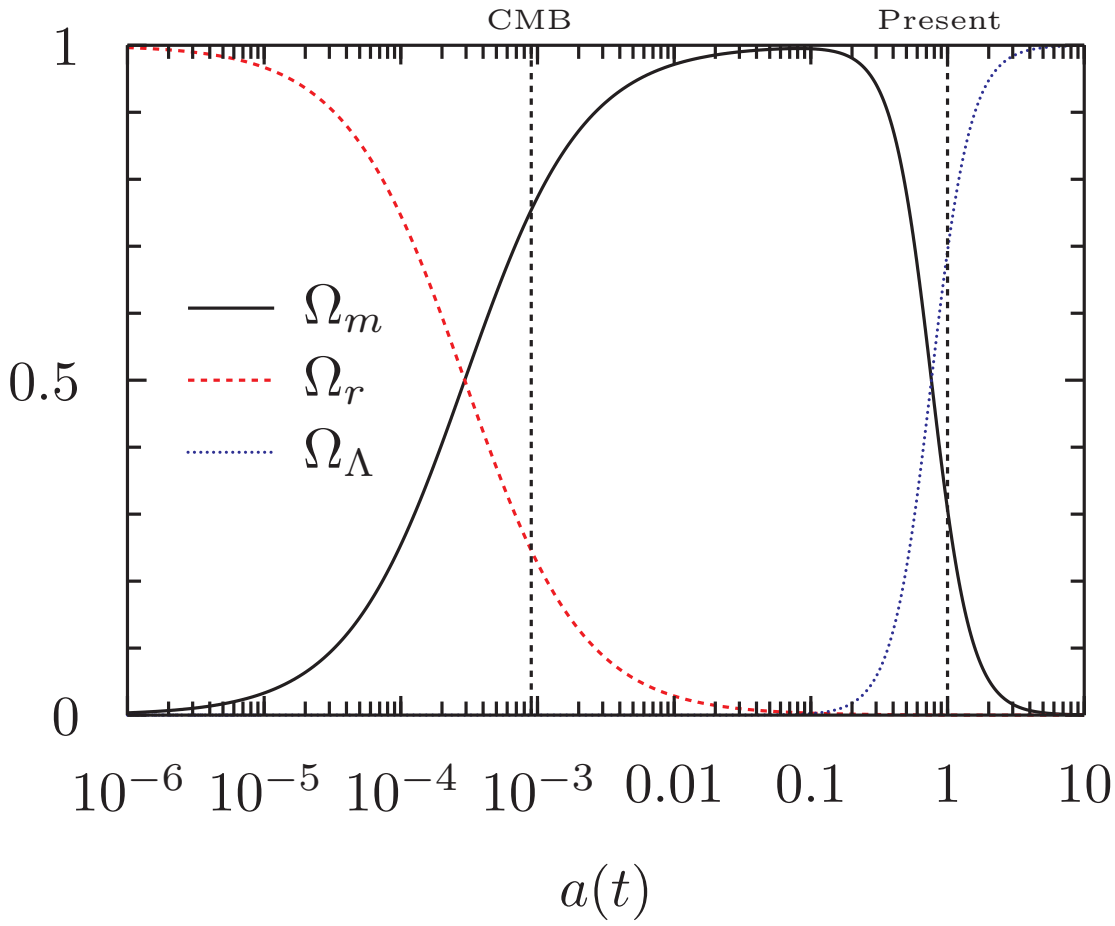


Figure 2.1: Evolution with scale factor, $a(t)$, of the density parameters describing the main components of the universe. The present day and the CMB are marked with dotted lines at $a(t) = 1$ ($z = 0$) and $a(t) = 9.2 \times 10^{-4}$ ($z = 1090.43 \pm 0.54$) respectively.

2.2 Structure formation

If you can look into the seeds of time
And say which grain will grow and which will not
Speak, then, to me

– William Shakespeare (Banquo, *Macbeth*, Act 1 Scene 3)

We have described the current best fit model of cosmology. However, all discussion thus far has involved a universe that is entirely homogenous and isotropic. This is only the case *on average* over scales that span many hundreds of Mpc. On scales smaller than this the universe is filled with a wealth of structure, in the form of galaxies in groups, clusters, filaments and walls that span many Mpc, and huge voids with low densities containing few galaxies. We illustrate this structure in Figures 2.2 and 2.3. Figure 2.2 is a projection of a density field from the Jubilee N-body simulation. The image is 500 h^{-1}Mpc across, 20 h^{-1}Mpc deep and is taken from the $z = 0$ output from the simulation. We show in Figure 2.3 an observational example of this structure from the Sloan Digital Sky Survey (SDSS). Each dot on the image represents a galaxy. Clearly, there is a gap that needs to be bridged between an entirely smooth FLRW universe and the universe we see around us today. In this section we consider the theoretical underpinnings of that bridge, in the form of linear structure formation. In the final section we describe how numerical simulations can greatly aid our modelling of the structure formation process.

2.2.1 Overview

The modern view of how structure formation proceeded is as follows. In the early universe there was an epoch of *inflation* (Guth, 1981), during which the scale factor increased exponentially. During this phase of very rapid expansion it is thought that very small quantum fluctuations in the field driving inflation (there may have been more than one field) were increased in size dramatically, such that they covered a range of scales from the very large ($\lambda > d_H$) to the very small ($\lambda \ll d_H$). This inflationary epoch ended in a process known as *reheating*, during which the energy in the field driving inflation was converted into the primordial material contents of the universe. At this point the universe was incredibly hot and dense and inertia from the inflationary expansion resulted in it continuing to expand. This has been termed the 'big bang'.

At this stage, shortly after the big bang, two points should be made. Firstly, the

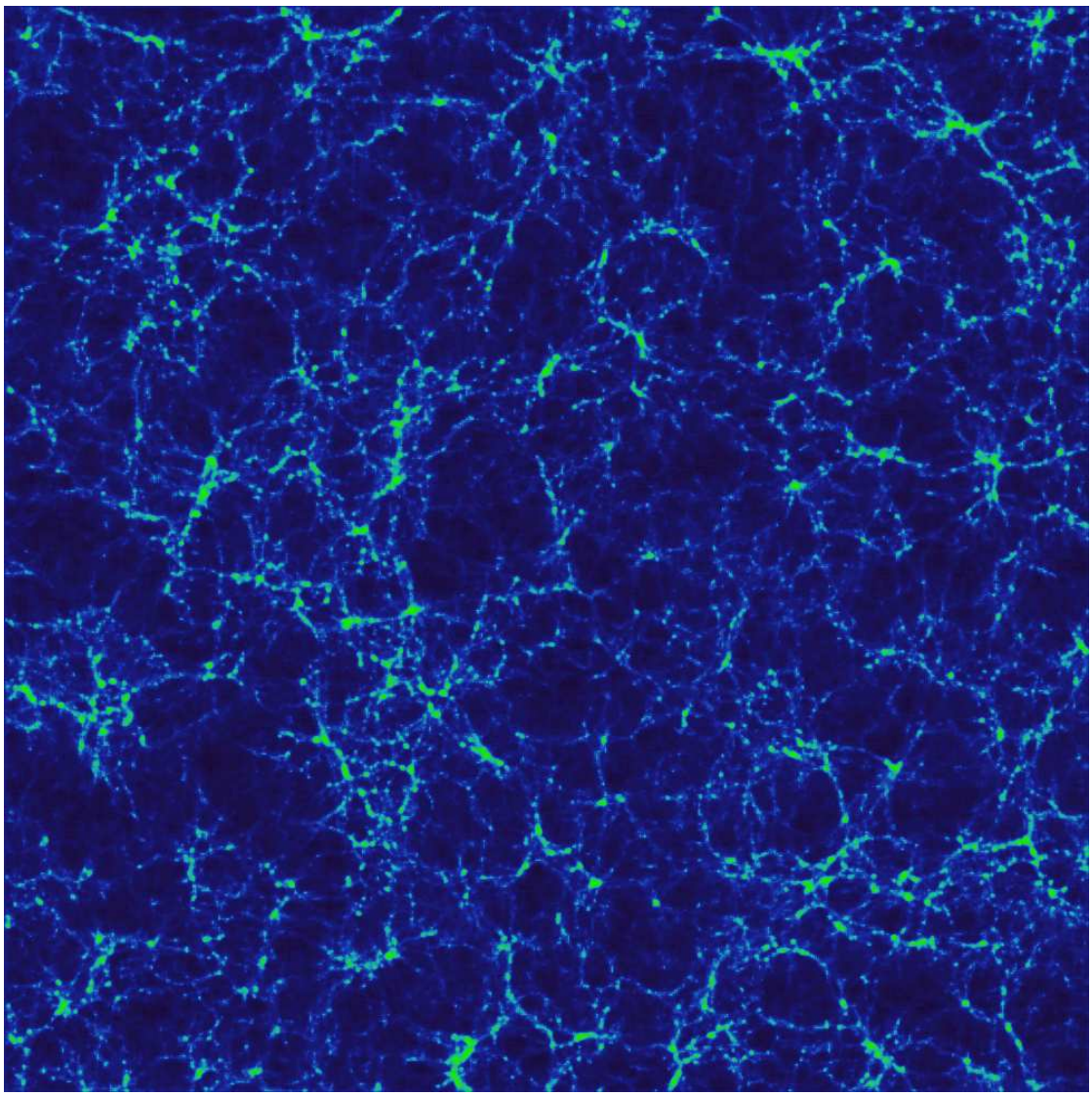


Figure 2.2: Structure in the Jubilee simulation at $z = 0$. The image is a projection of the density field in the simulation in a slab 500 h^{-1}Mpc a side, and 20 h^{-1}Mpc deep.

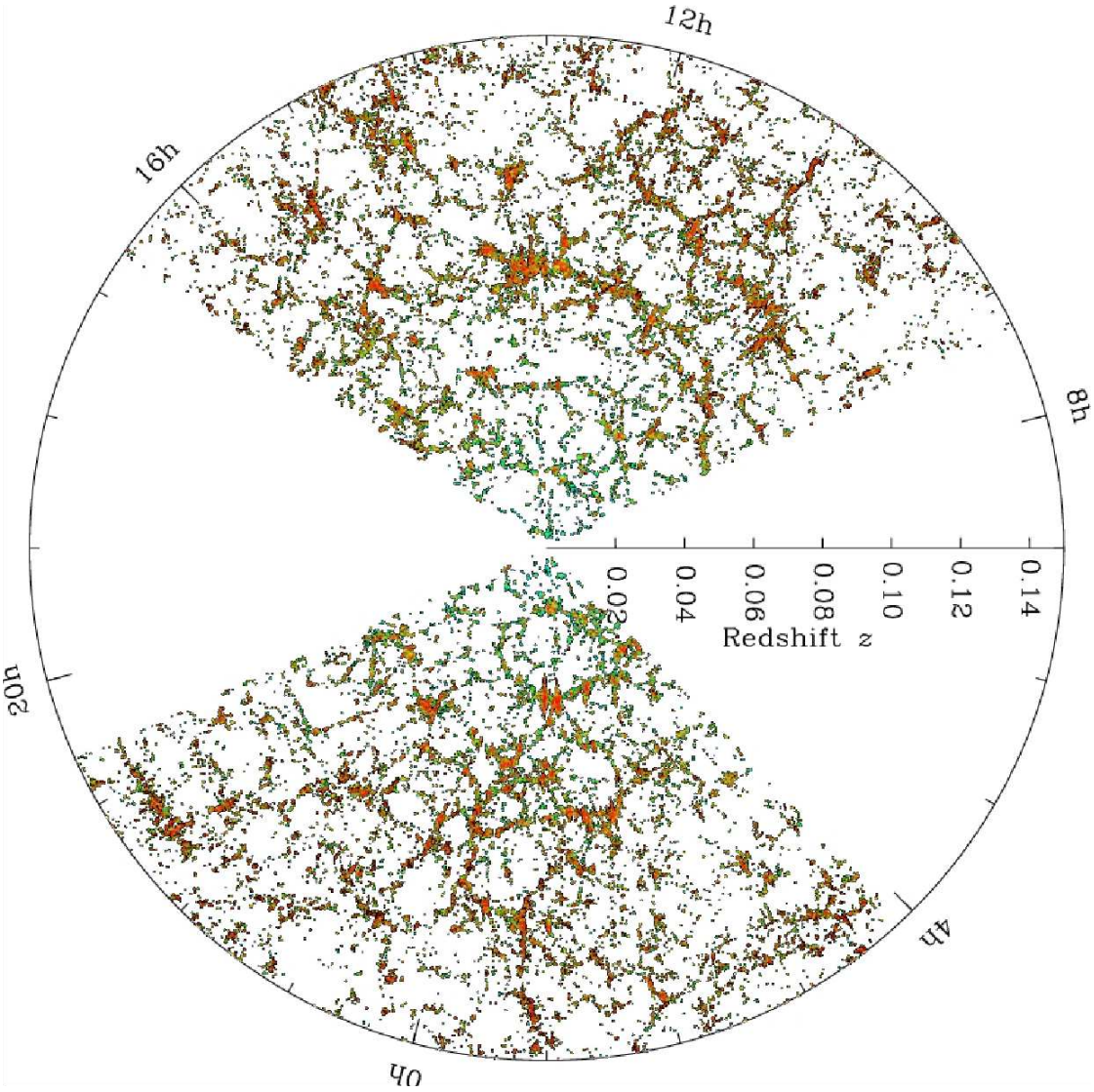


Figure 2.3: Structure in the Sloan Digital Sky Survey. The image used galaxies that were within -1.25 to 1.25 degrees of elevation. Each dot represents a galaxy. The galaxies are coloured according to the ages of their stars, with the redder points showing galaxies that are made of older stars *Credit: M. Blanton and the Sloan Digital Sky Survey.*

dominant component of the universe was relativistic particles. Secondly, a spectrum of density perturbations existed in the primordial universe that were created by quantum fluctuations from the inflationary epoch. These density perturbations are thought to be the seeds of the structure we see in the universe today. Eventually, the radiation domination gave way to matter domination, at $z_{eq} = 3391 \pm 60$ (Planck Collaboration, 2013c). 'Matter' in this context consists of baryons and dark matter. As dark matter has a low-to-zero interaction cross-section with baryons and photons (Bergström, 2012) and the baryons couple to photons strongly, it is necessary to treat these two fluids separately when considering their perturbations in the early universe. The baryons and photons coupled together until the time of *recombination*, $z_{rec} = 1090.43 \pm 0.54$ (Planck Collaboration, 2013c), when the universe had cooled to the extent that atoms of hydrogen, helium and lithium could form stably, free from harassment by ionising, high-energy photons. This was the point from which the photons from the primordial universe could stream freely across the entire universe and these are the photons that reach us today in the form of the CMB.

The coupling of the photons and baryons up to recombination created a marked signature in the CMB: the BAOs. These were oscillations that occurred due to the infall under gravity of baryons into potential wells (created by the density perturbations from inflation and seeded mainly by dark matter) and their subsequent outflow due to the build-up of radiation pressure as the baryon-photon fluid in the regions became denser. The dark matter participated in this only via gravity. Because of this the BAO signature in the distribution of dark matter at early times is minimal but grows over cosmic time such that it is clearly imprinted on the distribution of dark matter today.

Throughout the radiation- and early-matter-dominated epochs it is necessary to describe the physics of the universe using the full structure of the theory of General Relativity. This is for two reasons. One, the universe was radiation-dominated, or close to it, so was affected strongly by photons – photons, having zero rest mass, require a relativistic description – and two, density perturbations exist on scales larger than the Hubble distance ($\lambda > d_H$). The effect of perturbations entering the Hubble volume (defined as a sphere with radius d_H) as it expands needs a full relativistic treatment. Despite this, much of the theory of structure formation that we require in this work can be couched in the Newtonian limit, i.e. we can neglect considerations that arise due to relativistic effects and use results from the Newtonian theory of gravity.

The main tool used in this thesis is the N-body simulation, which simulates cosmolo-

gical volumes using Newtonian gravity. The use of the Newtonian limit in these simulations is reasonable, as the outputs from cosmological N-body simulations involve structures that lie well within the Hubble distance and well after radiation domination. Since we present results that, in certain cases, are based on very large volume N-body simulations (the largest being the Jubilee simulation with a box length of 6 h^{-1}Gpc, which compares to the present day Hubble distance of $d_H \sim 9.8\ h^{-1}$Gpc), there is a question as to whether relativistic effects might invalidate the Newtonian limit. This is not something to be taken lightly and has been the topic of recent debate, spurred on by the advent of simulations that span the Hubble length. The consensus is that, even on these large scales, N-body simulations give an accurate view of structure formation (Chisari & Zaldarriaga, 2011; Haugg et al., 2012; Green & Wald, 2012).

2.2.2 Linear perturbation theory in the Newtonian limit

Given an expanding universe with small initial perturbations we now address, using the Newtonian limit, how structures can grow. In the main we follow Padmanabhan (1993). To model the growth of structures that form under gravitational collapse we consider two classic fluid equations: the equation for mass conservation and the Euler equation; along with the Poisson equation:

$$\frac{\partial \rho}{\partial t} + \nabla(\rho \mathbf{v}) = 0, \tag{2.37}$$

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}.\nabla)\mathbf{v} + \frac{1}{\rho}\nabla p + \nabla\phi = 0, \tag{2.38}$$

$$\nabla^2 \phi = 4\pi G \rho. \tag{2.39}$$

The approach is to examine what the effect would be on these equations of taking small perturbations to the densities, pressures and velocities, of the form $\rho = \rho_b + \delta\rho$ etc. Before we do this we note that for the background, unperturbed densities and velocities: a) the background pressure is zero (in the Newtonian limit), b) the densities and velocities do not vary spatially (due to the FLRW background space-time) and c) the velocity at any position is attributable to the Hubble flow, i.e. $\mathbf{v}(\mathbf{x}, t) = a(t)\mathbf{x}$.

We can combine equations 2.37–2.39 by taking the divergence of the Euler equation (equation 2.38). It is helpful to define the *linear overdensity*:

$$\delta(\mathbf{x}, t) \equiv \frac{\rho(\mathbf{x}, t) - \rho_b(t)}{\rho_b(t)}, \tag{2.40}$$

where ρ_b is the average background density. The following results are then appropriate *only* in the regime where $\delta < 1$, as higher-than-leading order terms in the expansion of the quantities in equations 2.37-2.39 will begin to become important past this point. After linearising the equations we find (Padmanabhan, 1993):

$$\ddot{\delta} + 2H\dot{\delta} = \left(4\pi G\rho_b + \frac{v^2}{a^2}\nabla^2\right)\delta, \tag{2.41}$$

where $v^2 = \delta p/\delta\rho$ is the velocity dispersion which relates the perturbed pressure with the perturbed density and we have expressed all spatially varying quantities using comoving coordinates:

$$\mathbf{x}'(t) = \frac{\mathbf{x}}{a(t)} \tag{2.42}$$

These coordinates are the common choice in N-body cosmological simulations. Equation 2.41 describes the growth of density perturbations in a FLRW background in the Newtonian limit. The physical picture represented by equation 2.41 is that of a damped oscillator with the term proportional to δ (the right-hand-side) containing a balance between gravitational attraction and pressure support. The damping term, $2H\dot{\delta}$, arises due to the background expansion and is termed 'Hubble drag.' Gravitational instabilities such as these were studied in the early 20$^{\text{th}}$ century for the case of 'spherical nebulae' collapsing under gravity (Jeans, 1902). The essential result from this work is that there exists a balancing point between the pressure term ($(v^2/a^2)\nabla^2\delta$) and the gravitation term ($4\pi G\rho_b\delta$) on the right-hand-side of equation 2.41. The balancing point can be described by either the Jeans length, R_J, which is the length-scale beyond which gravity will prevail over the pressure term, or the Jeans mass, M_J, which is the mass enclosed by a sphere of radius R_J:

$$R_J = c_s\sqrt{\frac{\pi}{G\rho_b}}, \tag{2.43}$$

$$M_J = \frac{4\pi}{3}\rho R_J^3, \tag{2.44}$$

where c_s is the sound speed in the fluid.

We now look at the particular case of perturbations dominated by a pressureless fluid, for example dark matter. In this case the right-hand-side of equation 2.41 contains only a gravitation term and the equation as a whole contains only derivatives with respect to time. The equation then has a solution with two modes: i) a decaying mode with $\delta \propto t^{-1} \propto a^{-3/2}$; ii) a growing mode with $\delta \propto t^{2/3} \propto a$. This pressureless scenario was approximately true for overdensities in the universe from shortly after matter domination through to the recent epoch of dark-energy domination (Figure 2.1), and is referred to as the Einstein-de Sitter (EdS) universe. We note the result that in the dark matter-dominated regime the growing mode quickly became the important mode with time, such that the overdensities grew self-similarly in the linear regime for much of the history of the universe.

For the case of a universe dominated by a cosmological constant the right-hand-side of equation 2.41 goes to zero. This is because, a) the cosmological constant has no pressure perturbations (note this is not true for certain dark-energy scenarios) and b) the background matter density, ρ_b, goes to zero as the universe expands at an ever-accelerating rate. We see in this case that the growing mode becomes frozen: overdensities no longer grow. This is the scenario that will soon be the case for our universe.

We have described some of the more important results of linear perturbation theory. We now turn to the situation of overdensities that grow past $\delta \sim 1$.

2.2.3 Non-linear regime: the spherical collapse model

In general, when an overdensity has grown past $\delta \sim 1$ it is very difficult to describe its properties with much precision analytically. This is one of the motivations behind N-body simulations, which attempt to solve the problem of structure formation in the non-linear regime. Despite this, in cases where there exists a high-level of symmetry, such as for a collapsing spherical overdensity, we can make some headway. We briefly describe here the spherical collapse model of Gunn & Gott (1972), as in § 2.5 we will be using this model to help identify dark matter haloes in our simulations.

For a spherical overdensity collapsing in a background density field the equation of

motion for the radius of the overdensity is

$$\ddot{R} = -\frac{GM}{R^2} = -\frac{4\pi G}{3}\rho_b(1+\bar{\delta})R, \tag{2.45}$$

where $\bar{\delta}$ is the average overdensity in the sphere. This is derivable using Newtonian theory by considering a test particle at a radius, R, from the centre of the system. This particle will feel a net gravitational force towards the centre of the sphere due to all the mass contained within its radius. The mass outside of its radius, being spherically symmetric, gives no net contribution to the gravitational force on the particle (which is a statement of Newton's shell theorem).

Equation 2.45 has a parametric solution:

$$\frac{R}{R_t} = \frac{1}{2}(1-\cos\,\theta)\;;\quad \frac{t}{t_t} = \frac{1}{\pi}(\theta - \sin\,\theta), \tag{2.46}$$

where R_t is the maximum radius, which occurs at 'turnaround', specifically, at the turnaround time, t_t. The physical situation is then as follows. The overdense spherical region initially expands with the background but begins to slow in its expansion due to the attractive influence of its self-gravity. It expands until it reaches its turnaround point. After this point it collapses in a time-reversal of the expansion. At $R = 0$, which occurs at the collapse time of $t_c = 2t_t$, the overdensity is assumed to virialise. This results in the matter within the overdensity obeying the virial relation linking the kinetic energy (K.E.) and potential energy (P.E.) of the system: $2\,\text{K.E.} + \text{P.E.} = 0$. At this point the overdensity of the region is well past $\delta \sim 1$.

We can now match this parameterised solution for the top-hat collapse to the linear theory prediction, which obviously will break down a short time after the overdensity begins growing. For overdensities growing as $\delta \propto a \propto t^{2/3}$, i.e. for the growing mode in a pressureless fluid such as dark matter, the density contrast at turnaround is (Padmanabhan, 1993): $\bar{\Delta}_t = \bar{\rho}_t/\rho_b = (3\pi/4)^2 \approx 5.55$. By comparison the linear prediction for the overdensity at this time is $\bar{\delta}_t = \frac{3}{5}(3\pi/4)^{2/3} \approx 1.06$. At collapse we have $\bar{\Delta}_c = 18\pi^2 \approx 178$, and $\bar{\delta}_c = \frac{3}{5}(3\pi/2)^{2/3} \approx 1.686$. Note that the overdensities referred to are the average overdensities in the spherical region. When referring to these quantities in future we will drop the bar from their symbols.

Somewhat confusingly, the standard terminology in the literature refers to $\Delta = \rho/\rho_b$ *and* $\delta = \rho/\rho_b - 1$ both as the 'overdensity', despite the fact that these quantities differ by

one. In this work we use an uppercase Δ to refer to the former and a lowercase δ to refer to the latter (which is typically considered to be in the linear regime, although this is not the case in Chapter 5). Both Δ and δ are employed extensively throughout this thesis; Δ in particular, as we use this quantity to define dark matter haloes in our simulations (§ 2.5).

2.3 The power spectrum

Here we briefly outline the concept of the power spectrum of a density field because this is an important statistical description for some of the results we shall be considering later. In § 2.4.6 we will also use the concept of a power spectrum to describe how initial conditions are set for N-body simulations.

A power spectrum describes the variance of fluctuations in a density field at different scales (wavenumbers). If the power spectrum is larger for a particular wavenumber, k, then fluctuations with a scale of $\lambda \sim 2\pi/k$ will typically have a greater amplitude. Consider an overdensity field:

$$\delta(\mathbf{x}) = \rho(\mathbf{x})/\rho_b - 1. \tag{2.47}$$

This can be expressed as a sum of Fourier modes in a finite cubical box, such as a simulation volume with length L:

$$\delta(\mathbf{x}) = \frac{1}{L^3} \sum_n \delta_n \, e^{-i\mathbf{k}_n.\mathbf{x}}, \tag{2.48}$$

where the $\mathbf{k}_n$s are wavenumbers forming a cubic lattice with spacing $2\pi/L$. In the limit of $L \to \infty$, i.e. for an infinite volume, we have

$$\delta(\mathbf{x}) = \frac{1}{2\pi^3} \int \delta(\mathbf{k}) \, e^{-i\mathbf{k}.\mathbf{x}} \, d^3x. \tag{2.49}$$

We then define the power spectrum, $P(k)$, as

$$\langle |\delta_n|^2 \rangle = P(k), \tag{2.50}$$

where $k = |\mathbf{k}|$. For a field that is statistically isotropic and homogenous the power spectrum completely describes the statistics of the system. We can use the power spectrum to describe the variance of a field as a function of some smoothing scale, R:

$$\sigma^2(R, z) = \frac{1}{2\pi^2} \int_0^\infty k^2 P(k) W^2(k; R) \mathrm{d}k, \tag{2.51}$$

where $W(k, R)$ is the window function that describes the type of smoothing. For smoothing on scales with a radius of $R = 8h^{-1}$Mpc, this relation gives the σ_8 cosmological parameter, which is the variance in the density field of the universe at scales of 8 h^{-1}Mpc. Historically, this parameter was thought to have a value of around 1 (for example see Padmanabhan, 1993) and it is currently measured to be 0.826 ± 0.012 (based on the 'Planck + WP + highL + BAO' dataset of Planck Collaboration, 2013c). The σ_8 parameter is one of the six parameters mentioned in § 1.2 that describe the universe in the ΛCDM model.

It is often useful to consider the *dimensionless power spectrum*[5]:

$$\mathcal{P}(k) \equiv \frac{k^3}{2\pi^2} P(k), \tag{2.52}$$

which is defined so that scale-invariant fluctuations have $\mathcal{P} = \text{const}$.

In Chapter 5 we will be concerned with fluctuations that have been projected onto the inside surface of a sphere, which is how we observe many quantities on the sky, for example the CMB radiation. This projection reduces the dimensionality of the density fields and so we recast the power spectrum into a spherical, 2-dimensional form. The equivalent expression to equation 2.47 is

$$\frac{\Delta T(\theta, \phi)}{T} = \frac{T(\theta, \phi) - \bar{T}}{\bar{T}}, \tag{2.53}$$

where we have used the symbol 'T' since we are often concerned with temperature fluctuations on the sky (for example the variation in the temperature of the CMB induced by the ISW effect). The Fourier expansion of equation 2.53 now requires the use of spherical harmonics, $Y_{\ell m}(\theta, \phi)$, such that the spherical, 2-dimensional form of equation 2.48 is

$$\frac{\Delta T(\theta, \phi)}{T} = \sum_{\ell=2}^{\infty} \sum_{m=-\ell}^{+\ell} a_{\ell m} Y_{\ell m}(\theta, \phi). \tag{2.54}$$

[5]This is sometimes referred to as $\Delta^2(k)$. We avoid using this terminology and use $\mathcal{P}(k)$, to avoid any confusion with the overdensity, Δ.

The power spectrum, C_ℓ, is then defined via the $a_{\ell m}$ components in a manner analogous to equation 2.50:

$$\langle |a_{\ell m}|^2 \rangle = C_\ell. \tag{2.55}$$

C_ℓ has no dependence on the angular coordinates, (θ, ϕ) due to rotational invariance. In general $\ell \sim \pi/\theta$ rad^{-1}.

2.4 N-body simulations

I could be bounded in a nutshell and count myself a king of infinite space, were it not that I have bad dreams.

– William Shakespeare (Hamlet, Act 2 Scene 2)

Structure formation in the universe itself, when the density perturbations enter the non-linear regime of $\delta > 1$, becomes rather complex. Over the last 50 years or so it has become standard to attempt to model this process using computer simulations. In an ideal world we would be able to describe the evolution of the density fields of the components that make up the universe (dark matter, baryons, radiation and dark energy) in an arbitrarily precise manner. This is an unrealisable goal as in any simulation the spatial resolution is governed by two factors: 1) how large a patch of the universe we are simulating, and 2) how many tracer particles we use to follow the evolution of the density field. These two factors push in opposite directions: the larger the simulation volume the better, as we more accurately capture the overall statistical properties of the universe; the higher the number of particles the better as we resolve finer details of the density and velocity fields. Simulations therefore always represent a compromise between what length-scales in the universe we want to investigate versus how finely we want to resolve the structure forming on those scales.

In this thesis we look solely at gravity-only N-body simulations. Much work in the field of cosmological simulations today is focussed on trying to accurately model the interaction of baryons on top of the underlying dark matter field. In the simulations we use here details of the baryonic physics are ignored and we model the baryonic component of the universe only via its gravitational interactions.

2.4.1 History of N-body simulations

The first N-body simulation was a remarkable experiment by Holmberg (1941). This was a computerless simulation where lightbulbs (74 of them) were used to represent particles in two interacting galaxies on a two dimensional plane. The net gravitational force on each particle was calculated by measuring the received flux of light at the positions of each particle, as this quantity, like gravity, has an inverse-square dependence. The first computer-based N-body simulations were conducted by von Hoerner (1960) for $N = 16$, and shortly after by Aarseth (1963) ($N = 100$). Since these early efforts the growth of particle counts in simulations has roughly followed Moore's law and doubled every two years. The first simulations applied specifically to the question of the growth of structure in cosmology were undertaken by Press & Schechter (1974) and Haggerty & Janin (1974). Today simulations are run that contain tens to hundreds of billions of particles and cover cosmological volumes up to the Hubble volume. As the field has progressed a number of advances have taken place in order to meet specific challenges to the N-body approach. For recent reviews of the current status of N-body simulations see Kuhlen et al. (2012) and Dehnen & Read (2011). For an older review that covers the historical progress of cosmological simulations see Bertschinger (1998).

2.4.2 The collisionless Boltzmann equation

To fully model the non-linear regime of structure formation we need to solve, for the particles that make up the density field of the universe, the *Boltzmann equation*:

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \dot{\mathbf{x}}\frac{\partial f}{\partial \mathbf{x}} - \frac{\partial \Phi}{\partial x}\frac{\partial f}{\partial \dot{\mathbf{x}}} = \Gamma(f), \tag{2.56}$$

where $f(\mathbf{x}, \dot{\mathbf{x}}, t)$ is the *distribution function*; $\Phi(\mathbf{x}, t)$ is the gravitational potential; and $\Gamma(f)$ is the *encounter operator*. The distribution function is the probability per unit phase-space volume of a number of particles occupying an infinitesimally-sized volume in phase space, such that: $dN = f(\mathbf{x}, \dot{\mathbf{x}}, t)\ d^3\mathbf{x}d^3\dot{\mathbf{x}}$. The encounter operator, $\Gamma(f)$, describes close interactions, i.e. collisions, in the system and is, in general, a complicated, non-trivial function (Dehnen & Read, 2011). It is important in systems where the relaxation time is short compared to the lifetime of the system. This is the case in certain stellar systems, for example the centres of galaxies or globular clusters, where we have $t_{system} \gg t_{relax}$ and the system is termed *collisional*. For $t_{system} \ll t_{relax}$ the encounter operator is not important and can be set to $\Gamma(f) = 0$. These systems are described as *collisionless*. The

distinction is of relevance to cosmological simulations because the relaxation time of the universe is many times longer than its current age. On cosmological scales then, we model a collisionless system, with the encounter operator set to zero:

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \dot{\mathbf{x}}\frac{\partial f}{\partial \mathbf{x}} - \frac{\partial \Phi}{\partial x}\frac{\partial f}{\partial \dot{\mathbf{x}}} = 0. \tag{2.57}$$

This markedly simplifies the approach that needs to be taken to solve the problem. Despite this, equation 2.57 is still a non-linear partial differential equation in seven dimensions[6]. It is currently not possible to solve 2.57 analytically for an arbitrary dynamical system. This is in part due to the fact that the distribution function develops stronger gradients over time (Dehnen & Read, 2011). However, the equation is analytically useful because, if one takes moments of the equation, one can derive the Jeans equations (Jeans, 1915) which describe the motions of ensembles of particles, for example stars.

The N-body simulation solution to equation 2.57 is to use a Monte-Carlo approach by randomly sampling the distribution function at some early time in the universe's history, using N particles that represent the density field. The particles are then evolved under the influence of gravity to a later period in the universe's history. This approach captures many of the fundamental aspects of structure formation and is able to produce results that match our observations with good accuracy.

2.4.3 Configuration

A cosmological N-body simulation will typically consider a fixed comoving volume of the universe. This makes the comoving coordinates of equation 2.42 a natural choice to describe the particle positions. The time variable is often recast and there are a number of choices that can be made (typically based on how the time-stepping in the simulation is implemented). A common choice is *conformal time*:

$$d\eta \equiv \frac{dt}{a(t)}. \tag{2.58}$$

The equations that then need to be solved to evolve the simulation – the force calculation and the Poisson equation – become:

[6] In contrast to the collisional version of the equation which is a 6N-dimensional equation, where N is the number of particles under consideration. For the collisionless case correlations between particles in the system are ignored.

$$\frac{d^2\mathbf{x}_i'}{d\eta^2} + \frac{\dot{a}}{a}\frac{d\mathbf{x}_i'}{d\eta} = -\nabla\Phi(\mathbf{x}', \eta), \tag{2.59}$$

$$\nabla^2\Phi(\mathbf{x}', \eta) = 4\pi G a^2[\rho(\mathbf{x}', \eta) - \rho_b]. \tag{2.60}$$

2.4.4 Force calculation

The biggest issue faced by simulations is that of undertaking an appropriate force calculation. A naive pair-wise calculation of the gravitational force on each particle in a simulation takes the form:

$$\mathbf{F}_{ij} = \sum_{i\neq j}^{N} G m_i m_j \frac{\mathbf{x}_j - \mathbf{x}_i}{|\mathbf{x}_j - \mathbf{x}_i|^3}. \tag{2.61}$$

This equation runs into an immediate problem in that when particles are close together the force between them becomes very large. Divergences of this sort must be avoided in numerical calculations, so the force is softened for small distances, for example using a Plummer sphere potential:

$$\mathbf{F}_{ij} = \sum_{i\neq j}^{N} G m_i m_j \frac{\mathbf{x}_j - \mathbf{x}_i}{(|\mathbf{x}_j - \mathbf{x}_i|^2 + \epsilon^2)^{3/2}}. \tag{2.62}$$

After accounting for the force softening[7], another issue that has to be overcome is the scaling with time of the force calculation. For a pairwise particle-particle (PP) scheme this is of order $O(N^2)$, taking $10N^2 - N$ operations (Hockney & Eastwood, 1988). For an increasing number of particles, even with a parallel algorithm using many independent computational cores, this scaling effectively rules out the PP approach to solving collisionless simulations with large numbers of particles. Instead, the force calculation is typically taken as an approximation on larger scales and a more exact calculation on shorter ones. There are a variety of algorithms that implement this approach (see for example Dehnen & Read (2011)). One of the earliest developed was the particle-particle-particle-mesh (P^3M) method (Efstathiou, 1979), which calculates the short-range forces

[7]There are multiple softening schemes that are employed. See for example Dyer & Ip (1993); Athanassoula et al. (2000); Dehnen (2001).

between particles that are near to each other and the long-range forces by smoothing the particles onto a density grid and calculating the force on each particle from this grid via the Poisson equation (equation 2.60). This is the approach adopted in the CUBEP^{3}M code (Harnois-Deraps et al., 2012), which is the code used to run the simulations on which this is work is based. Other examples include particle-mesh-only (PM) codes, tree-based codes, multipole-expansion codes and adaptive grid-based codes.

2.4.5 Computational requirements

As the size of simulations has grown so too has the quality of the infrastructure on which they are run. Today's largest simulations, consisting of almost a trillion particles (Kuhlen et al., 2012), require huge resources both in terms of data storage and processor memory and performance. To be able to keep pace with Moore's law, N-body codes have made the transition to parallel architecture and today's codes, for example CUBEP^{3}M, are typically hybrid codes that utilise both OpenMP and MPI parallel implementations. In addition the large amounts of data produced by these simulations can run into the hundreds of terabytes and in the near future petabyte datasets will become standard.

2.4.6 Initial conditions

As N-body simulations are essentially deterministic in nature, the two ingredients they require are a set of physical laws and a set of initial conditions. These are typically set at an early time, $z \sim 100$, using linear theory to calculate the locations of the N particles in a manner that represents statistically the density field of the early universe. For a detailed description of how this is achieved see Martel (2005).

From the power spectrum of CMB temperature anisotropies we can derive a 3-dimensional power spectrum that describes linear fluctuations in the matter field at a given redshift. This is achieved using linear theory and a Boltzmann code (for example Bertschinger (1995), Seljak & Zaldarriaga (1996) or Lewis et al. (2000)). An important aspect of setting initial conditions is therefore choosing an appropriate initial redshift where density fluctuations are in the linear regime. We discuss this and other criteria for the choice of starting redshift for a simulation below.

From the initial power spectrum of the matter field we assign a Gaussian random number to the amplitude of each of the modes in the power spectrum. These amplitudes are then used to distribute the particles in the simulation box. The boxsize limits the longest wavelength mode that can be contained in the simulation volume and the interparticle

spacing sets a limit on the shortest (via the Nyquist-Shannon sampling theorem).

To translate from the density field in Fourier space to the actual displacement of particles in the box from a regular grid (or glass), it is typical to use the Zel'dovich approximation (Zel'dovich, 1970), which gives the displacement for the particles as given by first order Lagrangian perturbation theory (Moutarde et al., 1991; Buchert, 1993) (hereafter 1LPT), i.e. based on the gradient of the density field. This is the approach adopted for the CUBEP^{3}M runs in this work.

The simulations considered in this thesis are all modelled using particles representing the matter field that interact only via gravity. This is despite the fact that the density field contains a contribution from both baryons and dark matter (we neglect the contribution from radiation as this is very small in the late-time universe, we also assume that no inhomogeneities exist in the dark energy field), the former interacting via complicated physical processes in the actual universe. Modelling the entire matter field in this way results in the signature of the baryons – the BAOs – being present in the density field in the simulations.

2.4.6.1 Criteria for setting initial conditions

There is debate regarding the suitability of the Zel'dovich approximation in setting initial conditions (Crocce et al., 2006, 2010; Lukic et al., 2007; Tinker et al., 2008; Knebe et al., 2009; Reed et al., 2012). Care must be taken when setting a choice for the initial redshift to ensure that artefacts (transients) are minimised. A detailed study by Lukic et al. (2007) emphasises the need for a suitable choice for the initial redshift when applied to a particle-mesh (PM) code (MC^2 – see Heitmann et al., 2005; Heitmann et al., 2008) using 1LPT. Reed et al. (2012) recommend the use of second-order Lagrangian perturbation theory (2LPT) in setting initial conditions, based on their study using two tree-based codes: GADGET-2 (Springel, 2005) and PKDGRAV (Stadel, 2001).

Lukic et al. (2007) proposed two criteria to guide the choice of initial redshift for simulations employing 1LPT: (1) whether the amplitude of the initial power spectrum modes in the box are in the linear regime, and (2) whether the initial particle displacement is small enough so that the particle grid distortion is relatively small. The first criterion sets a minimum for the initial redshift, based on the requirement that the dimensionless power spectrum, $\mathcal{P}(k) = k^3P(k)/2\pi^2$, be less than some arbitrary, small value at the initial redshift (Lukic et al. (2007) used $\mathcal{P}(k)_{max} \leq 0.01$). As smaller boxes for a given particle number probe higher values of k, earlier initial redshifts are required for them. Similarly,

for a fixed box size, simulations with greater numbers of particles also require earlier initial redshifts. The second criterion places a more stringent bound on the initial redshift. The example discussed in Lukic et al. (2007) was that the particles were displaced, on average, by no more than $\delta_{in}^{rms} = 0.3\Delta_p$, where $\Delta_p = L_{box}/n_p$ is the interparticle spacing and δ_{in}^{rms} is the root-mean-squared displacement of the particles.

Finally there is a numerical consideration, which is that the particle displacements must be adequately captured by the precision of the variables being used to define the particle positions. If the displacements are too low, which can occur if the initial redshift is too high, then numerical noise will distort the initial conditions.

2.4.7 The CUBEP^{3}M code

The CUBEP^{3}M code calculates short-range direct particle-particle forces combined with a long-range PM force calculation, making it a P^{3}M (particle-particle-particle-mesh) code. It is massively parallel and runs efficiently on either distributed- or shared-memory machines. This is achieved via a cubical, equal-volume domain decomposition combined with a hybrid OpenMP and MPI approach. CUBEP^{3}M scales well up to thousands of processors and to date has been run on up to 21,976 computing cores (Iliev et al., 2008, 2010; Harnois-Deraps et al., 2012). We note that such large simulations yield very large amounts of data, with just the particle data from the Jubilee simulation of 6000^3 particles amounting to 4.7 TB per time-slice, providing significant challenges in the data handling and analysis.

2.4.8 CUBEP^{3}M runs used in this work

The simulations undertaken in this work are summarised in Table 2.2. They follow from 3072^3 (29 billion) to 6000^3 (216 billion) particles in a wide range of box sizes from 11.4 h^{-1}Mpc up to 6 h^{-1}Gpc. Spatial resolutions range from 0.18 h^{-1}kpc to 50 h^{-1}kpc while particle masses range from 3.6×10^3 $h^{-1}\mathrm{M}_\odot$ to 7.5×10^{10} $h^{-1}\mathrm{M}_\odot$. This allows dark matter haloes to be resolved in a very large mass range, with a low end of 7.3×10^4 $h^{-1}\mathrm{M}_\odot$ (for a 20 particle halo) in our smallest volumes, and no upper limit in the halo mass, since our largest volumes approach the size of the observable universe.

The CUBEP^{3}M code uses first-order Lagrangian perturbation theory (1LPT), i.e. the Zel'dovich approximation (Zel'dovich, 1970), to place particles in their initial positions. The initial redshift when this step takes place is given for each simulation in table 2.2, along with the values of $\mathcal{P}(k)_{max}$ and $\delta_{in}^{rms}/\Delta_p$ for our simulations (see § 2.4.6). We see that for all of our simulations both of these criteria are satisfied.

Table 2.2: N-body simulation parameters for the runs used in this work. Background cosmology is based on the WMAP 5-year results. Minimum halo mass and the redshift of formation of the first halo are based on a 20 particle minimum. Resolution here refers specifically to the softening length (see Harnois-Deraps et al. (2012) for details). See § 2.4.6 for details regarding the quantities $\mathcal{P}(k)_{max}$ and $\delta_{in}^{rms}/\Delta_p$.

Box Size h^{-1}Mpc	N_{part}	Mesh	Resolution h^{-1}kpc	$m_{particle}$ $h^{-1}M_\odot$	$M_{halo,min}$ $h^{-1}M_\odot$	$z_{\rm in}$	$z_{\rm final}$	$\mathcal{P}(k)_{max}$	$\delta_{in}^{rms}/\Delta_p$	$z_{\rm firsthalo}$
11.4	3072^3	6144^3	0.18	3.63×10^3	7.63×10^4	300	8	2.0×10^{-5}	0.037	41
20	5488^3	10976^3	0.18	3.63×10^3	7.63×10^4	300	8	2.0×10^{-5}	0.045	44
114	3072^3	6144^3	1.86	3.83×10^6	7.63×10^7	300	6	1.2×10^{-5}	0.069	30
425	5488^3	10976^3	3.87	3.69×10^7	7.35×10^8	300	3	9.5×10^{-6}	0.068	25
1000	3456^3	6912^3	14.47	1.96×10^9	3.85×10^{10}	150	0	4.6×10^{-4}	0.057	17
3072	3072^3	6144^3	50.00	7.49×10^{11}	1.49×10^{12}	100	0	2.8×10^{-5}	0.051	11
3200	4000^3	8000^3	40.00	4.06×10^{10}	8.12×10^{11}	120	0	4.5×10^{-4}	0.052	11
6000	6000^3	12000^3	50.00	7.49×10^{10}	1.50×10^{12}	100	0	2.8×10^{-5}	0.051	11

2.4.9 Cosmology

We base our simulations on the 5-year WMAP results (Dunkley et al., 2009; Komatsu et al., 2009). With the exception of two of our runs (the 1 h^{-1}Gpc and 3.2 h^{-1}Gpc boxes), the cosmology used for the simulations was the 'Union' combination from Komatsu et al. (2009), based on results from WMAP, baryonic acoustic oscillations and high-redshift supernovae; i.e. $\Omega_m = 0.27$, $\Omega_\Lambda = 0.73$, $h = 0.705$, $\Omega_b = 0.044$, $\sigma_8 = 0.8$, $n_s = 0.96$. The 1 h^{-1}Gpc and 3.2 h^{-1}Gpc boxes were based on the slightly different 'Alternative' combination from the same paper; i.e. $\Omega_m = 0.279$; $\Omega_\Lambda = 0.721$, $h = 0.701$, $\Omega_b = 0.046$, $\sigma_8 = 0.817$, $n_s = 0.96$. These results are consistent with the recent cosmology results of the Planck collaboration (Planck Collaboration, 2013c), where, considering a combination of data from Plank, WMAP and LSS surveys, the parameters were calculated to be: $\Omega_m = 0.307$, $\Omega_\Lambda = 0.691$, $h = 0.678$, $\Omega_b = 0.048$, $\sigma_8 = 0.829$ and $n_s = 0.961$. The power spectrum and transfer function used for setting initial conditions were generated using CAMB (Lewis et al., 2000).

2.5 Halo finding

2.5.1 Motivation

Structure formation proceeds via peaks in the linear density field of the early universe growing into highly non-linear structures via gravitational attraction. The growth has been termed 'hierarchical' as we find that low density peaks grow via accretion of material in their local environments and by a series of mergers with other density peaks. A popular formalism for describing this process is the identification of the non-linear, high density peaks as dark matter haloes. These haloes, appropriately defined, are concomitant with virialised regions in the universe where galaxies reside. The hierarchical growth of structure in this picture can then be viewed as a series of halo mergers, commencing with very small haloes at early times and continuing all the way to the present, with haloes growing larger via both mergers and the accretion of smaller haloes. In cosmological simulations it is usual to attempt to track the progress of the growth of structure by identifying haloes within the simulation volume across a range of redshifts.

2.5.2 Halo finding approaches

There are two main approaches to finding haloes in cosmological simulations. The first is the spherical overdensity (SO) method of Lacey & Cole (1993). In this method haloes

are assumed to be spherical. The extent of a halo is governed by a free parameter, the overdensity criterion, Δ, which is a cut-off in density with respect to some background density (typically either the matter background density or the critical density of the universe). Δ here is the equivalent to the average overdensity, $\bar{\Delta}$, from the spherical collapse model (§ 2.2.3). In the SO algorithm, spheres are grown from a central location until the enclosed overdensity of the sphere is equal to the overdensity criterion. The main differences between halofinding codes that use this approach lie in how the centres of the candidate haloes are identified. The mass enclosed in these spherical haloes is given by

$$\mathrm{M}_{\Delta} = \frac{4\pi\Delta\rho_m}{3}\mathrm{R}_{\Delta}^3. \tag{2.63}$$

The second approach is the Friends-of-Friends (FOF) algorithm. This algorithm is also based on one parameter, b, the linking length parameter. The algorithm finds haloes that contain particles that are within $b\Delta_p$ of at least one other particle in the halo. While the SO algorithm produces a halo that is by definition spherical, the FOF algorithm creates haloes that are arbitrarily shaped. For a recent comparison project on halofinding codes see Knebe et al. (2011).

We employ three halo-finding codes in our analysis in this thesis: CubeP^{3}M's own on-the-fly SO halofinder (hereafter 'CPMSO') (Harnois-Deraps et al., 2012), the Amiga Halo Finder (hereafter 'AHF') (Gill et al., 2004; Knollmann & Knebe, 2009) and the FOF halofinder from the Gadget-3 N-Body cosmological code (an update to the publicly available Gadget-2 code (Springel, 2005)).

The CPMSO halofinder utilises a fine mesh from the CubeP^{3}M code (a mesh with spacing of $\Delta_p/2$) to identify local peaks in the density field. The code first builds the fine-mesh density using either Cloud-In-Cell (CIC) or Nearest-Grid-Point (NGP) interpolation. It then proceeds to search for and record all local density maxima above a certain threshold (typically set to 100 above the mean density) within the physical volume. It then uses quadratic interpolation on the density field to determine more precisely the location of the maximum within the densest cell. The halo centre determined this way agrees closely with the centre-of-mass of the halo particles. Each of the halo candidates is inspected independently, starting with the highest peak. The grid mass is accumulated in spherical shells of fine grid cells surrounding the maximum until the mean overdensity within the halo drops below Δ. While the mass is accumulated it is removed from the mesh, so that no mass element is double-counted. This method is inappropriate for finding sub-haloes

as within this framework they are naturally incorporated in their host haloes. Because the haloes are found on a grid of finite-sized cells and spherical shells constructed from them, it is possible, especially for the low-mass haloes, to overshoot the target overdensity. When this occurs we use an analytical halo density profile to correct the halo mass and radius to the values corresponding to the target overdensity. This analytical density profile is given by the Truncated Isothermal Sphere (TIS) profile (Shapiro et al., 1999; Iliev & Shapiro, 2001) for overdensities above ~ 130 and $1/r^2$ for lower overdensities. The TIS density profile has a similar outer slope (the relevant one here) to the Navarro, Frenk and White (NFW) profile (Navarro et al., 1997) but extends to lower overdensities and matches well the virialisation shock position given by the Bertschinger self-similar collapse solution (Bertschinger, 1985). For further details on the CPMSO method see Harnois-Deraps et al. (2012).

Details for AHF can be found in Knollmann & Knebe (2009). The algorithm identifies density peaks using a nested set of grids that are finer-grained in regions of higher density. Haloes are then identified by collecting particles together that are contained within isodensity contours on each level of the grid hierarchy. This allows AHF to identify sub-haloes within host haloes as the algorithm progresses from high resolution grids to lower resolution ones. AHF then removes particles that are unbound and recalculates halo properties based on the remaining bound particles.

The specifics of the FOF halofinder packaged in with the Gadget-3 code have not been detailed currently in any publication but the algorithm itself is outlined in Davis et al. (1985). The main difference in the algorithm that exists in the Gadget-3 version is that the code is parallelised for distributed-memory machines. Specifically, haloes are found in local subvolumes of the simulation assigned to individual MPI tasks (created using the Gadget-3 domain decomposition which utilises a space-filling Peano-Hilbert curve – for details see the Gadget-2 paper Springel (2005)) and then haloes that extend spatially beyond the edges of the subvolumes are linked together in a final MPI communication step. We have altered the Gadget-3 code to read CubeP^{3}M's particle output format and significantly reduced its memory footprint by stripping away extraneous data structures.

Finally, due to limitations in the scaling of the codes with processor numbers, large memory footprint and incompatible data structures, in order to apply the AHF and FOF algorithms to our data, it was necessary to split the simulation time-slices into a number of subvolumes and run the halofinding algorithms on each subvolume independently. Each subvolume included a buffer zone which overlapped with the neighbouring ones for correct

handling of haloes straddling two or more sub-regions. We then stitched the subvolumes back together to create the final AHF and FOF halo catalogues, removing any duplicated structures in the overlapping buffers. Although somewhat more expensive than applying each halo finder directly on the full data, this approach allows the handling of much larger amounts of data than otherwise possible and provides additional flexibility in terms of computational resources needed for post-processing. This was a major component in the successful completion of this thesis.

Chapter 3

The halo mass function

In this chapter we investigate how the dark matter halo mass function – the abundance of haloes by mass – evolves with redshift. Our halo catalogue data span a redshift range of $z = 0 - 30$, allowing us to probe the mass function from the Dark Ages to the present. Much of the content of this chapter can be found in Watson et al. (2012), which has been accepted for publication in MNRAS.

3.1 Introduction

As noted in (§ 1.2.1), the halo mass function is difficult to accurately pin-down using observations. The two main issues faced by observers in measuring the mass function are 1) building large enough datasets of galaxies or clusters to reduce the statistical uncertainties that arise from low number counts and 2) relating the distribution of observed light to the underlying distribution of mass. In the case of galaxies the first problem is becoming much less of an issue as we perform larger and larger surveys (Driver et al., 2011; Coil et al., 2011; LSST Dark Energy Science Collaboration, 2012; Dawson et al., 2013). The latter problem is overcome via studying the bias of galaxies as a tracer of the underlying matter peaks, a topic that has been well studied (see More (2011) for a discussion of how accurate our current understanding is). For the case of clusters, both of the above observational issues represent major challenges. The number of observed clusters remains low despite recent, ongoing and future surveys (Varela et al., 2009; Vikhlinin et al., 2009; Rozo et al., 2010; Balaguera-Antolínez et al., 2011; Takey et al., 2011; Gilbank et al., 2011; Pillepich et al., 2012; Mehrtens et al., 2012) and cluster counting via the Sunyaev-Zel'dovich (SZ) effect (Fowler et al., 2010; Planck Collaboration, 2011a; Hasselfield et al., 2013; Planck Collaboration, 2013d). Relating cluster observables to their underlying masses is a well-

studied problem in cosmology (Vikhlinin et al., 2009; Planck Collaboration, 2011c,b; Rozo et al., 2012b,a; Angulo et al., 2012; Planck Collaboration, 2013d). The existence of an intrinsic scatter in the relationships between observables and masses highlights that trying to obtain an accurate measurement of the high-mass tail of the mass function is a stiff observational challenge.

In contrast, the mass function is relatively easy to probe via N-body cosmological simulations (see Lukic et al. (2007) for a review of older work; more recent studies include Cohn & White (2008); Tinker et al. (2008); Crocce et al. (2010); Courtin et al. (2011); Bhattacharya et al. (2011); Angulo et al. (2012)) and can, to an extent, be understood and modelled through analytic arguments (Press & Schechter, 1974; Bond et al., 1991; Lee & Shandarin, 1998; Sheth & Tormen, 2002; Maggiore & Riotto, 2010; Corasaniti & Achitouv, 2011; Lim & Lee, 2012).

Much has been been written on the topic of the mass function by groups working with N-body simulations. To date the majority of this work has looked at the mass function at low redshift ($z \lesssim 2$) using haloes derived using the Friends-of-Friends (FOF) algorithm (Davis et al., 1985). Despite this there has been a significant amount of work investigating the mass function at high redshift (Heitmann et al., 2006; Reed et al., 2007; Lukic et al., 2007; Cohn & White, 2008; Klypin et al., 2011) and various authors have studied the mass function derived via the alternative Spherical Overdensity (SO) algorithm (Lacey & Cole, 1994). For example Tinker et al. (2008) calculate a fitted mass function for SO haloes, valid for $z \lesssim 2$; Jenkins et al. (2001) employed both the FOF and SO algorithms finding similar results for each for $z < 5$ (although the SO results had more scatter across redshifts and cosmologies, see the discussion on universality below); Cohn & White (2008) found good agreement between the FOF and SO mass functions at $z = 10$ for appropriately chosen algorithm parameters (see below and § 3.4.4.2); Reed et al. (2007) focussed on FOF haloes but compared the results to SO haloes finding the number density of haloes to be systematically lower in the case of SO haloes. Fitted functions are abundant in the literature (see § 3.4.2), with the majority being calibrated using FOF haloes. There is currently no robust mass function available for haloes derived using the SO algorithm at redshifts higher than $z \sim 2$, which represents a gap in the literature that we address in this work. In this chapter we add our own fits for both FOF and SO halo mass functions and exhibit their differences across a range of masses and redshifts.

There are shortcomings of both halofinding algorithms of which the reader should be aware. This is reflective of the fact that there is no specific definition of a dark matter

halo that is agreed upon in the literature. The FOF algorithm identifies regions that are bounded by constant density contours in real space (see § 3.4.4.1) whereas the SO algorithm creates an artificial, spherically-bounded region. Both of these outputs are referred to as 'haloes'. In reality, a dark matter halo is never perfectly spherical and at higher masses and redshifts the virialisation process that occurs in high-density peaks is often incomplete, creating a tension between objects identified via the SO algorithm and physical reality. In the FOF case there is a systematic effect inherent to the algorithm known as 'overlinking' which occurs when two haloes are linked together by a bridge of particles. When this occurs the resultant objects are not physically interpretable as virialised collapsed density peaks, rather they are a representation of a complex system that is undergoing relaxation or merging. Previous authors (Davis et al., 1985; Ma & Bertschinger, 1995; Cole & Lacey, 1996; Lukic et al., 2009) have estimated that 15-20% of all FOF haloes calculated with the standard linking length of 0.2 are objects that have been bridged together in this manner. However, it is important to note that for a suitable choice of linking length it is possible to obviate the effect of overlinking: in general the lower the linking length parameter the lower the amount of overlinking. The correspondence between SO haloes and FOF haloes has been studied, for example in White (2001) and White (2002), and, recently, empirical relations have been calculated by Lukic et al. (2009), Courtin et al. (2011) and More et al. (2011) that relate SO and FOF haloes, in terms of their masses. This is looked at in more detail in § 3.4.4.1 where we discuss the differences between the two halo types and the effect this has on the mass functions derived from them. In this work we make no attempt to quantify the various underlying effects that lead to different mass function results when derived via the two algorithms. The effects, outlined in § 3.2.1 and discussed in detail in § 3.4.4, are numerous and their interplay complex.

Jenkins et al. (2001) published results suggesting that the mass function was perhaps 'universal', i.e. independent of redshift or cosmology, when expressed in suitable units. A careful study by White (2002) (building on earlier results on defining halo mass (White, 2001)) showed that deviations from a universal form were small but existent. Other negative results have been published by various authors (Reed et al., 2007; Tinker et al., 2008; Crocce et al., 2010; Courtin et al., 2011; Bhattacharya et al., 2011). It was originally noted by Jenkins et al. (2001) that in order to produce a universal mass function from FOF haloes one needs to take a constant linking length – the key parameter that the FOF algorithm uses, see § 2.5 – across all redshifts or cosmologies being considered. Tinker et al. (2008) observed a clearer departure from universality for SO haloes than for FOF

haloes (in line with the earlier result of Jenkins et al. (2001)), a fact that More et al. (2011) propose is due to the effect of taking a fixed overdensity criterion – the key parameter of the SO algorithm, see § 2.5 – rather than specific overdensity criteria bespoke to a given redshift or cosmology. One is therefore left with a choice: either adjust one's halo definitions to attempt to produce a universal mass function or keep one's halo definitions fixed and expect the function to deviate from a universal form. In this work we choose the latter. Deciding upon a suitable 'universal' redshift-dependent halo overdensity or linking length parameter is still an open question (although Courtin et al. (2011) and More et al. (2011) go some way towards providing a solution), due to complicating factors such as the triaxiality of haloes in the SO case – a purely spherical halo is an ideal that is never realised – and overlinking in the FOF case. We shy away from demanding a universal mass function and pin our halo definitions down firmly in § 2.5.

This chapter is laid out as follows. In § 3.2 we outline the definitions of what we are referring to by the term 'halo' and discuss various nuances that occur. In § 3.3 we present our mass function results and fits. In § 3.4 we present a summary of the work and a discussion of various points that arise. In this chapter we utilise all the simulations listed in Table 2.2, with the exception of the 3072^3, 3.072 h^{-1}Gpc run.

3.2 Halo definitions

In this chapter we compare results from all of the halofinders mentioned in § 2.5, i.e. the CPMSO, AHF and FOF halofinders. We choose the overdensity criteria for identification of spherical overdensity haloes to be Δ_{178}, i.e. an overdensity of 178 times the background matter density. This is a common choice, motivated by the top-hat model of non-linear collapse in an EdS universe that we described in § 2.2.3. The overdensity criterion is usually taken to be an overdensity with respect to either the background matter density or the background critical density. For an EdS universe, i.e. a universe containing only pressureless matter, the two are the same. For a ΛCDM universe, at late times, the two deviate from each other due to the increasing dark energy component that contributes to the critical density of the universe but not to the matter density (Figure 2.1). Care must be taken to compare like-for-like overdensities, especially during the Λ-dominated epoch (note that all references in this work to overdensities refer to those with respect to the background matter density). Any given non-linear overdensity criterion can be mapped onto a threshold value in the linear regime. For Δ_{178} this corresponds to a linear overdensity of $\delta_c = 1.686$ in an EdS universe, as described in § 2.2.3. In a ΛCDM universe

the value of δ_c evolves to 1.674 at $z = 0$ due to the influence of dark energy. In the case of the FOF halofinder we follow various previous authors (Jenkins et al., 2001; Reed et al., 2003, 2007; Crocce et al., 2010; Courtin et al., 2011; Angulo et al., 2012) and use a linking length of 0.2 for our analysis.

3.2.1 Halo mass redshift dependence

There are a number of factors that influence the growth of halo mass. They can be divided into 'physical' mass growth, i.e. the growth of the haloes via mergers and accretion, and, following the nomenclature of Diemer et al. (2012), 'pseudo' mass growth. For a recent study on the relative effects of physical and pseudo mass growth on the SO mass function between $z = 0-1$ see Diemer et al. (2012) and references therein. An important point from this study is that a significant fraction of SO halo mass growth between $z = 0-1$ is due to pseudo-evolution of halo mass. This is growth of halo mass, defined in the SO sense, due to the evolution of the background density of the universe. As the background matter density of the universe decreases with the universe's expansion, the radius of a given SO halo will grow as it requires a lesser enclosed physical density to meet the overdensity criterion of Δ times the background density. In the FOF case there also exists a pseudo mass growth: as the linking length is defined as a constant length in co-moving coordinates, its physical length increases as z decreases. The implication of this is that for a halo with a static profile the extent of an FOF halo will increase with time. Fakhouri & Ma (2010) split FOF halo growth into an accretion component and a diffuse component, the latter containing an element of pseudo growth. Currently, there is no equivalent study to the Diemer et al. (2012) work on pseudo mass growth in SO haloes for FOF haloes.

In addition to the pseudo evolution of halo mass there is the evolution of δ_c (and Δ) due to dark energy dominance at low redshifts. This has well-studied implications for the SO halo mass definition (Lahav et al., 1991; Lacey & Cole, 1993; Eke et al., 1996). Because of this, $\delta_c(z)$ is often used as a parameter in mass functions (Press & Schechter, 1974; Sheth & Tormen, 2002; Reed et al., 2003, 2007; Courtin et al., 2011). In this work we provide an SO mass function fit for haloes in an ΛCDM universe that is based on Δ_{178} and do not include $\delta_c(z)$ as a variable in the fit. This is motivated by the fact that we parameterise redshift evolution occurring due to dark energy, pseudo mass growth and systematic effects such as halo mergers (which occur at different rates in different epochs) and non-sphericity of haloes. These latter effects are not captured by the evolution of $\delta_c(z)$. We also note that more correctly modelling the baryonic component of haloes will

have an effect on the mass function (estimated by Cui et al. (2012) to increase the mass function by $\lesssim 3\%$ for a Δ_{200} halo definition) not captured in this work.

An issue that pseudo mass evolution raises is that any redshift evolution of the mass function (whether expressed in terms of mass or $\ln\sigma^{-1}$, see § 3.3) is due to an interplay between halo definition and the physical increase of halo mass with time. In order to link halo definitions to observations and compare our results to previous studies on the halo mass function, we adopt the common halo definitions above and attempt to capture redshift evolution based on them.

3.3 The halo mass function

The mass function can be expressed in a number of ways. Here, we use the multiplicity function, $f(\sigma, z)$, which represents the fraction of mass that has collapsed to form haloes per unit interval in $\ln\sigma^{-1}$. This definition uses σ^{-1} as a proxy for mass, where σ^2 is the variance of the linear density field, given by

$$\sigma^2(\mathrm{M}, z) = \frac{D^2(z)}{2\pi^2}\int_0^\infty k^2 P(k) W^2(k;\mathrm{M})\mathrm{d}k, \tag{3.1}$$

where $P(k)$ is the power spectrum of the linear density field, $W(k;\mathrm{M})$ is the Fourier-space representation of a real-space top-hat filter containing mass M – assuming that the top-hat sphere encloses a region that contains a mean density identical to that of the universe – and $D(z)$ is the growth factor, normalised to the unity at $z = 0$ (Heath, 1977). The radius in real-space of the filter can be set using an overdensity criterion, Δ, as per the SO algorithm. The relationship between σ and halo mass for our two WMAP-5-based cosmologies is shown in Figure 3.1.

We define the halo multiplicity function as (Jenkins et al., 2001):

$$f(\sigma, z) \equiv \frac{\mathrm{M}}{\rho_m(z)}\frac{dN(\mathrm{M}, z)}{d\ln\sigma^{-1}}, \tag{3.2}$$

where $N(\mathrm{M}, z)$ is the mass function proper, that is to say the number count of haloes with a mass less than M per unit volume, and ρ_m is the mean matter density of the universe.

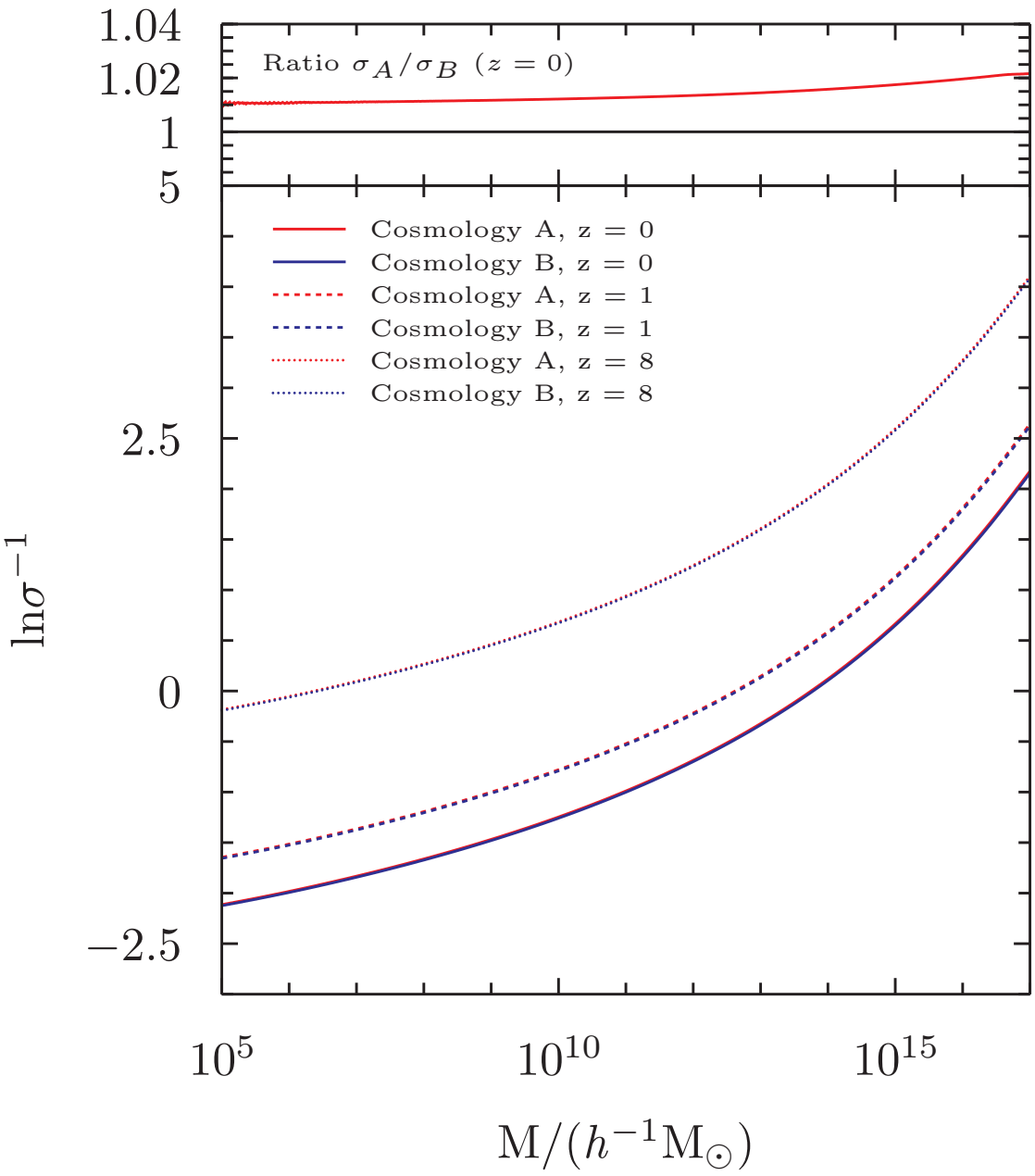


Figure 3.1: The relationship between the variance of the linear density field, σ, and mass for two ΛCDM cosmologies. 'Cosmology A' refers to the 'Union' model of Komatsu et al. (2009) with $\Omega_m = 0.27$, $\Omega_\Lambda = 0.73$, $h = 0.7$, $\Omega_b = 0.044$, $\sigma_8 = 0.8$, $n_s = 0.96$. 'Cosmology B' refers to the 'Alternative' model of Komatsu et al. (2009) with $\Omega_m = 0.279$; $\Omega_\Lambda = 0.721$, $h = 0.701$, $\Omega_b = 0.046$, $\sigma_8 = 0.817$, $n_s = 0.96$. The two cosmologies in question are very similar, with mass-σ relations that lie almost on top of one another. In the top panel of the figure we show the $z = 0$ difference between them as a ratio, showing that for the mass range 10^5 to 10^{17} $h^{-1}\mathrm{M}_\odot$ the difference is 1 to 2%.

3.3.1 Mass binning

The halofinders calculate values for the masses of all resolved haloes. It is then necessary, in order to construct $f(\sigma, z)$, to count the number of haloes in mass bins before converting the masses into values of $\ln\sigma^{-1}$. Rather than equation 2, for simulated halo catalogues we have

$$f(\sigma, z) \equiv \frac{\mathrm{M}^2}{\rho_m(z)} \frac{\Delta N_{sim}}{\Delta \mathrm{M}} \frac{d\ln \mathrm{M}}{d\ln\sigma^{-1}}, \tag{3.3}$$

where care must be taken in choosing the width of the bins in mass as this can potentially result in a source of systematic error. For a detailed evaluation of this error see Lukic et al. (2007), who note that the error is not significant as long as the bin widths, ΔlogM, do not exceed 0.5. We use bin widths for our analysis that remain constant in logM with ΔlogM=0.16. and assign the mass of a given bin to the average of all the haloes in the bin rather than the bin centre.

3.3.2 Error treatment

Throughout this chapter we use error bars as defined in Heinrich (2003):

$$\sigma_{\pm} = \sqrt{N_{haloes} + \frac{1}{4}} \pm \frac{1}{2}. \tag{3.4}$$

These error bars differ somewhat from the standard $1-\sigma$ Poissonian error bars (which are simply $\sigma_{\pm} = \pm\sqrt{N_{haloes}}$. They are defined based on the Pearson χ^2 statistic. For data described by a Poisson process with a mean μ, from which we observe n datapoints, the probability of observing this number, n, of haloes is given by

$$p(n, \mu) = \frac{e^{-\mu}\mu^n}{n!}, \tag{3.5}$$

and Pearson's χ^2 is

$$\chi^2(\mu, n) = \frac{(n-\mu)^2}{\mu}. \tag{3.6}$$

Having observed n events – in this case, having counted a number of haloes in a mass bin – we look to assign a central value for the mean in the bin. For our case we will be

fitting across a range of masses a function that describes the number of haloes in different mass bins. This function will be constructed by finding parameter values that minimise the χ^2 value across all the mass bins, weighted by the size of the error in each bin, i.e. by the uncertainty that the data we measure is describing reality accurately. With a mean assigned in each bin we then adopt a $1-\sigma$ interval for the error bars, which leads to Equation 3.4. The main advantages of using error bars in this form are: (1) for a bin with no haloes the error is still finite, (2) for a halo count of 1 the lower edge of the error bar does not reach zero. Note that for large numbers of haloes the errors tend to the expected $\sqrt{N}$ form. Since their initial adoption by Heitmann et al. (2006) and Lukic et al. (2007) the use of these error bars in mass function studies has become common.

3.3.3 Data treatment

In order to combine data across our simulations, certain systematic effects need to be accounted for. These include: accounting for the finite volumes of our simulations; adjusting the FOF haloes for a systematic overestimation in mass for haloes sampled with low particle numbers; the question of whether to remove or include sub-haloes from our AHF data and, finally, the choice of a lower limit to the number of particles a halo contains. The mass functions presented here are (1) corrected for finite volumes for simulations with a box size of less than 425 h^{-1}Mpc, (2) adjusted, in the case of our FOF haloes, to account for mass overestimation using the empirical correction of Warren et al. (2006), (3) constructed based only on host haloes (in the case of data from the AHF halofinder the host haloes have masses that include any sub-haloes they contain) and (4) have a minimum of 1000 particles in each halo. We now discuss each of these points in detail.

3.3.3.1 Finite volume correction

The volume of space modelled by a cosmological simulation is always finite. However, when we speak of the mass function we typically refer to a 'global' mass function, i.e. one that would correspond to an infinitely large volume. There is therefore a disconnect that needs to be bridged between simulations and an ideal, global mass function (Sirko, 2005; Bagla & Prasad, 2006; Power & Knebe, 2006). There exist a number of ways to address this issue. In this work we adopt an approach recently employed by Lukic et al. (2007) and Bhattacharya et al. (2011). We proceed by assuming that mass function universality holds strictly in the sense that the functional form is the same for both global and local volumes. This is similar to assuming that the mass function is universal across cosmologies

that differ in their mass-σ relations, as the effect of a finite box size is to set to zero the amplitude of any density fluctuations on a scale greater than the box size. There is evidence that the mass function is not universal (see § 3.4), that its functional form has a weak dependence on both redshift and cosmology. Therefore, any finite volume correction we make based on the assumption of universality is an approximation. We adopt the approach here despite this, as it is relatively straight-forward to apply and it brings our data into better agreement across different box sizes. We also briefly discuss below other possible methods that could be applied.

The approach is as follows. Assuming equation 3.2 refers to the mass function in an infinite simulation volume, we can re-write it as

$$\frac{dN}{d\mathrm{M}} = \frac{\rho_m}{\mathrm{M}^2} f(\sigma) \frac{d\ln\sigma^{-1}}{d\ln\mathrm{M}}, \tag{3.7}$$

and for a finite simulation volume we can write

$$\frac{dN'}{d\mathrm{M}'} = \frac{\rho_m}{\mathrm{M}'^2} f(\sigma) \frac{d\ln\sigma'^{-1}}{d\ln\mathrm{M}'}, \tag{3.8}$$

where $\sigma'(\mathrm{M}')$ is determined by the discrete power spectrum of the simulation in question. Also M' is a function of M defined such that $\sigma(\mathrm{M}) \equiv \sigma'(\mathrm{M}'(\mathrm{M}))$. The assumption that the mass function is universal allows us to say that the multiplicity functions, f, in equations 3.7 and 3.8 are identical. This then leads to the relation

$$dN = dN' \frac{\mathrm{M}'(\mathrm{M})}{\mathrm{M}}. \tag{3.9}$$

We next require some method of connecting $\sigma'(\mathrm{M}')$ and $\sigma(\mathrm{M})$. The extended Press-Schechter formalism (Bond et al., 1991) approximately connects $\sigma'(\mathrm{M}')$ and $\sigma(\mathrm{M})$ via

$$\sigma'(\mathrm{M}')^2 = \sigma(\mathrm{M})^2 - \sigma^2_{R(box)}, \tag{3.10}$$

where $\sigma^2_{R(box)}$ is the variance of fluctuations in spheres that contain the simulation volume. To align this to the simulations, we make another approximation and equate the spherical volume to the cubical simulation volume. The resulting mass-σ relations at $z = 0$ for the 114 h^{-1}Mpc and 20 h^{-1}Mpc boxes are shown in Figure 3.2.

The steps in the volume correction are as follows: calculate $\sigma^2_{R(box)}$ for the simulation box in question; for each bin in σ calculate an adjusted bin σ value, σ_{cor}, using $\sigma^2_{cor} = \sigma^2 - \sigma^2_{R(box)}$; adjust each halo count using equation 3.9, where M is the mass that corresponds to σ and M$'$ the mass that corresponds to σ_{cor}.

For comparative purposes we now discuss other solutions that have been employed to solve the finite volume issue. Reed et al. (2007) performed a number of different N-body simulations and calculated the range of σ for each one using the input power spectra of the simulations and the relationship between σ and mass for a finite box in the discrete case:

$$\sigma^2(\mathrm{M}, z) = D^2(z) \sum_{\mathbf{k}} |\delta_{\mathbf{k}}|^2 W^2(k; \mathrm{M}), \tag{3.11}$$

where $|\delta_{\mathbf{k}}|$ is the linear amplitude of the Fourier modes in the simulation at $z = 0$. This approach has the advantage of both correcting for finite volumes and also compensating for cosmic variance. One drawback to it is that each realisation has a different mass-σ relationship. We did not adopt this approach because it is a method more suited to multiple simulation runs, where the mass-σ relations can be averaged over to produce a mass function fit. We did not have the luxury of repeating our runs due to the large sizes of the simulations.

Bagla et al. (2009) choose to make no specific correction (although they remove any data points that are affected by more than a threshold level of error – 10% in the number counts – from their analysis, in a similar manner to Tinker et al. (2008)). This decision was motivated by the observation that the mass function is an unknown function that is deduced from simulation data so to place any a priori constraints upon it – such as a universal functional form for global and local mass-σ relations – is undesirable. This approach was applicable in the study undertaken by Bagla et al. (2009) because the box sizes used for their simulations were all relatively large ($\geq 256\ h^{-1}$Mpc). For our smaller boxes, not correcting for finite volumes would lead to undesirable systematic discontinuities across our simulations.

Yoshida et al. (2003) and Bagla & Prasad (2006) replaced equation 3.1 with

$$\sigma^2_{box}(\mathrm{M}, z) = \frac{D^2(z)}{2\pi^2} \int_{2\pi/L}^{\infty} k^2 P(k) W^2(k; \mathrm{M}) \mathrm{d}k, \tag{3.12}$$

which takes a cut-off in low k modes at the size of the box. Lukic et al. (2007) note

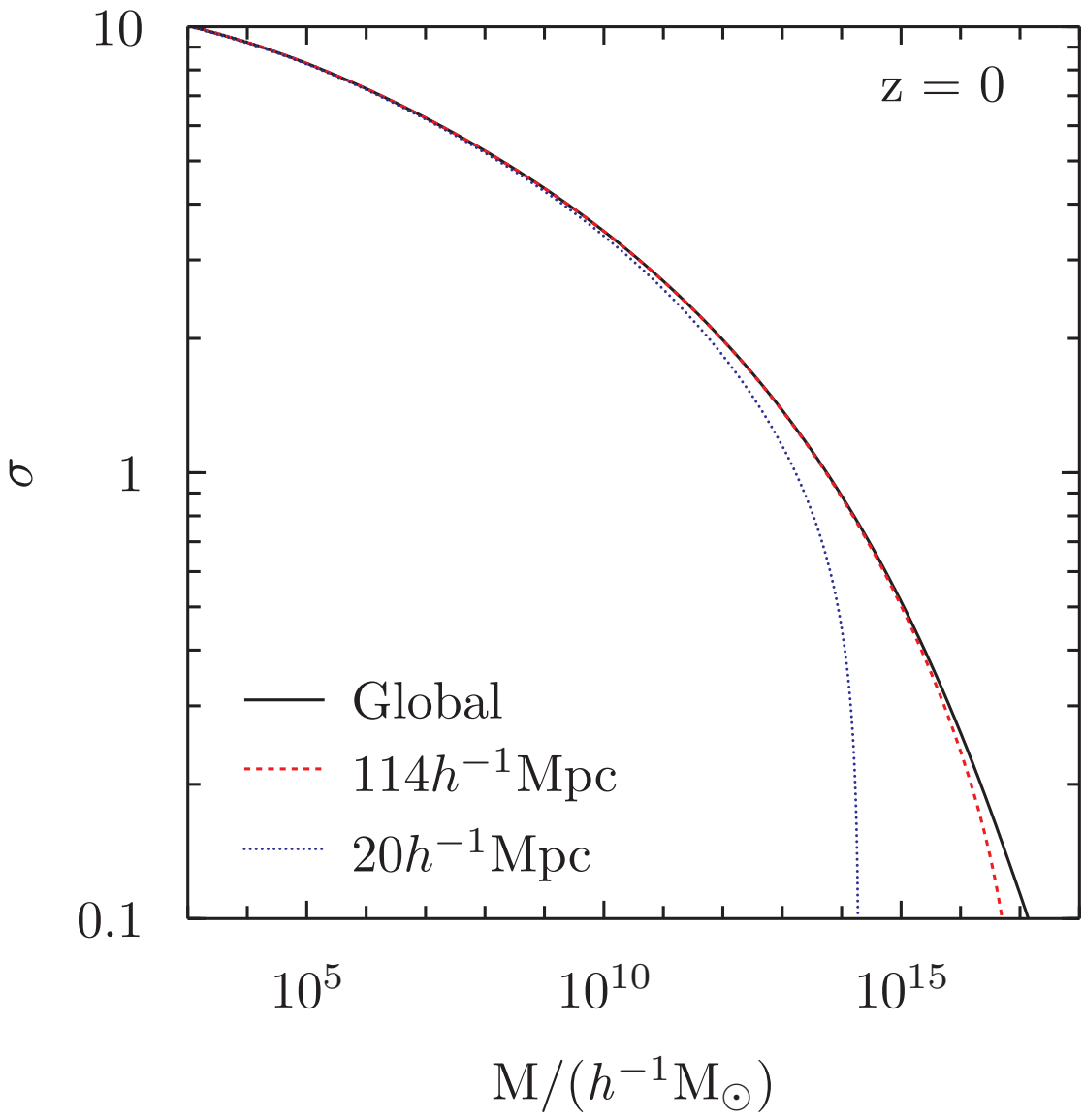


Figure 3.2: The finite volume correction to the mass function for small boxes. The global $M-\sigma$ relation is adjusted to give a relation based on the volume of the simulation box in question via equation 3.10. The (approximate) assumption of mass function universality is then employed to adjust the masses and numbers of dark matter haloes via equations 3.9 and 3.10.

that a correction of this form has a dependence on the accuracy of the mass functions at redshifts greater than $z = 5$. They also note that in comparison to their correction method it exhibits offsets and shape changes across different box sizes.

Finally, Barkana & Loeb (2004) utilised the extended Press-Schechter formalism to create a volume adjustment by taking the Sheth-Tormen mass function for a global volume and applying a correction to it to create a mass function suitable for a smaller volume. Since we are not adopting a mass function in the Sheth-Tormen form, this approach was inappropriate here.

3.3.3.2 Warren correction to FOF haloes

Warren et al. (2006) proposed a simple correction to the masses of FOF haloes that alleviates a systematic error in halo masses calculated via the FOF algorithm at low particle counts. This correction was devised based on analysis of FOF haloes at $z = 0$, but has been checked by Lukic et al. (2007) for FOF haloes at higher redshifts. The FOF algorithm overestimates the masses of haloes when there are low numbers of particles sampling the haloes. Warren et al. (2006) proposed the following correction to particle counts in haloes:

$$N_{corrected} = N(1 - N^{-0.6}). \tag{3.13}$$

We have adopted this correction in this work and all FOF data shown includes it. As we have also adopted a cut-off in particle number of 1000 the maximum effect this correction has on the masses of our FOF haloes is $\sim 2\%$.

There is debate regarding the appropriate correction to use. Bhattacharya et al. (2011) find that a correction of $N_{cor} = N(1 - N^{-0.65})$ is more suitable and Lukic et al. (2009) note that the halo concentration parameter for haloes with an NFW profile also affects the correction, a result that is corroborated by More et al. (2011). The more sophisticated corrections proposed by Lukic et al. (2009) and More et al. (2011) (the former provide a correction for FOF haloes with $b = 0.2$ whereas the latter provide a correction valid for different values of b) are based on individual halo concentrations. We have not adopted this approach as we do not possess profile information for our individual FOF haloes. The difference between the Warren correction and the More et al. (2011) correction (which can be as much as 15% for haloes with lower particle counts) becomes slight for haloes with many particles. For our smallest, 1000 particle haloes, the correction to the Warren

formula is of the order of $< 5\%$ depending on halo concentrations. Not accounting for this therefore introduces a small systematic error which in the very worst case of 1000 particle haloes would alter the FOF masses by $\sim 0.15\%$. In addition More et al. (2011) note that the Warren correction is specifically calculated to correct for the bias in the mass function itself whereas their study is one that corrects for the masses of individual FOF haloes when considered in isolation.

3.3.3.3 Sub-haloes

As noted in § 2.5 the AHF halofinder produces halo catalogues that contain sub-haloes – bound structures inside host haloes. For the purposes of constructing a mass function we have to make a choice as to how to deal with these haloes. The CPMSO and FOF halofinders both identify host haloes only, therefore for compatibility we make the natural choice of excluding the sub-haloes from the mass functions based on the AHF catalogue. The host haloes detected by AHF have properties, including mass, that are calculated based on all the bound matter contained within the halo, including the sub-haloes themselves. This results in the AHF host haloes being directly comparable to the CPMSO ones.

3.3.3.4 Low-end particle cutoff

We take a minimum particle cut-off of 1000 particles per halo in this study. This is conservative as often halo-finders report haloes with as little as 20 particles or even fewer. Lukic et al. (2007) recommend a 300 particle minimum for constructing a mass function with a pure PM code. Warren et al. (2006) take an 'aggressive' 400 particle minimum in their study and Tinker et al. (2008) a 'conservative' 400. It has been observed recently by Reed et al. (2012) that a minimum particle cut-off of at least 1000 is appropriate. Although this latter figure was quoted with tree-based N-body codes in mind – rather than a P^3M code – we use it to motivate our cut-off of 1000 particles. We note that for studies in other areas it might be acceptable to use haloes with lower particle counts, e.g. in clustering studies we need to reliably identify and locate a halo but not necessarily know its mass or other properties precisely. For mass function calculations it is important that haloes are assigned masses that are as close to their correct values as possible and more particles are required to ensure this is the case.

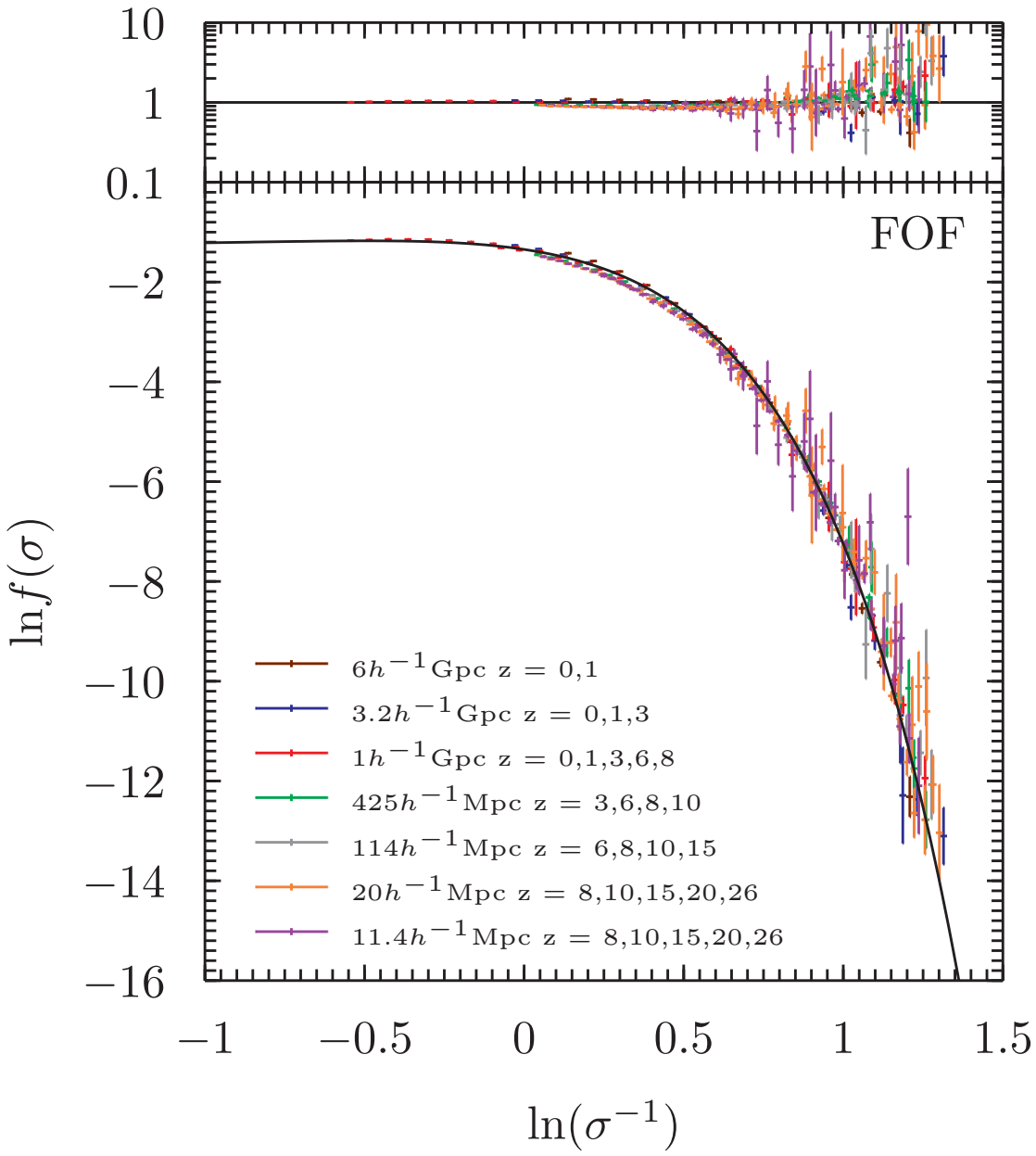


Figure 3.3: The FOF mass function across all simulations and redshifts with our fitted function (equation 3.14). The top panel shows the ratio of the data to the fitting formula.

3.3.4 The universal FOF function

We find that the halo mass function based on our FOF halo catalogues follows a universal fit applicable for data from all simulations across all redshifts. The fit lies within $\sim 10\%$ for most data points (Figure 3.3) and takes the following the form:

$$f(\sigma) = A\left[\left(\frac{\beta}{\sigma}\right)^{\alpha} + 1\right] e^{-\gamma/\sigma^2}, \tag{3.14}$$

where $A = 0.282$, $\alpha = 2.163$, $\beta = 1.406$, $\gamma = 1.210$. This fit is valid in the range: $-0.55 \leq \ln\sigma^{-1} < 1.31$, which at $z = 0$ corresponds to haloes with masses between 1.8×10^{12} and 7.0×10^{15} $h^{-1}\mathrm{M}_{\odot}$. Our largest halo, found in the 3.2 h^{-1}Gpc box at $z = 0$, has a mass of 6.4×10^{15} $h^{-1}\mathrm{M}_{\odot}$ and $\ln\sigma^{-1} = 1.17$, whereas our highest $\ln\sigma^{-1}$ value comes from a $z = 3$ halo in the same simulation, with a mass of 1.4×10^{14} $h^{-1}\mathrm{M}_{\odot}$. It should be noted that this extreme halo is most likely the result of the FOF algorithm linking together two large haloes via a bridge of particles – a systematic effect known as overlinking, discussed in § 3.4 below. At higher $\ln\sigma^{-1}$ values, i.e. for high mass/high redshift, rare haloes, the scatter about the fit increases dramatically due to shot noise.

Whilst it is a remarkable result that over such a large range of z and σ we observe a halo mass function that approximately conforms to a universal shape, we urge caution as the mass function is not completely universal and exhibits a modest redshift evolution (see § 3.4 for details).

3.3.5 Mass function for spherical overdensity haloes

3.3.5.1 Redshift evolution

We show in Figure 3.4 the mass function from our AHF haloes. In contrast to the FOF mass function it is clear that a universal fit is not appropriate for spherical overdensity haloes (a fact also borne out in the results from the CPMSO halofinder. For a comparison between the CPMSO and AHF halo mass functions see Figure 3.7 and discussion below). We again adopt the fitting function in equation 3.14 but with a parameterisation that includes a redshift-dependence, as discussed below. Since running AHF on multiple checkpoints from our largest simulations is computationally expensive, we utilise the data from the CPMSO on-the-fly halofinder in order to investigate the redshift-dependence of the parameters for the spherical overdensity case. We then re-scale this fit to match the AHF results.

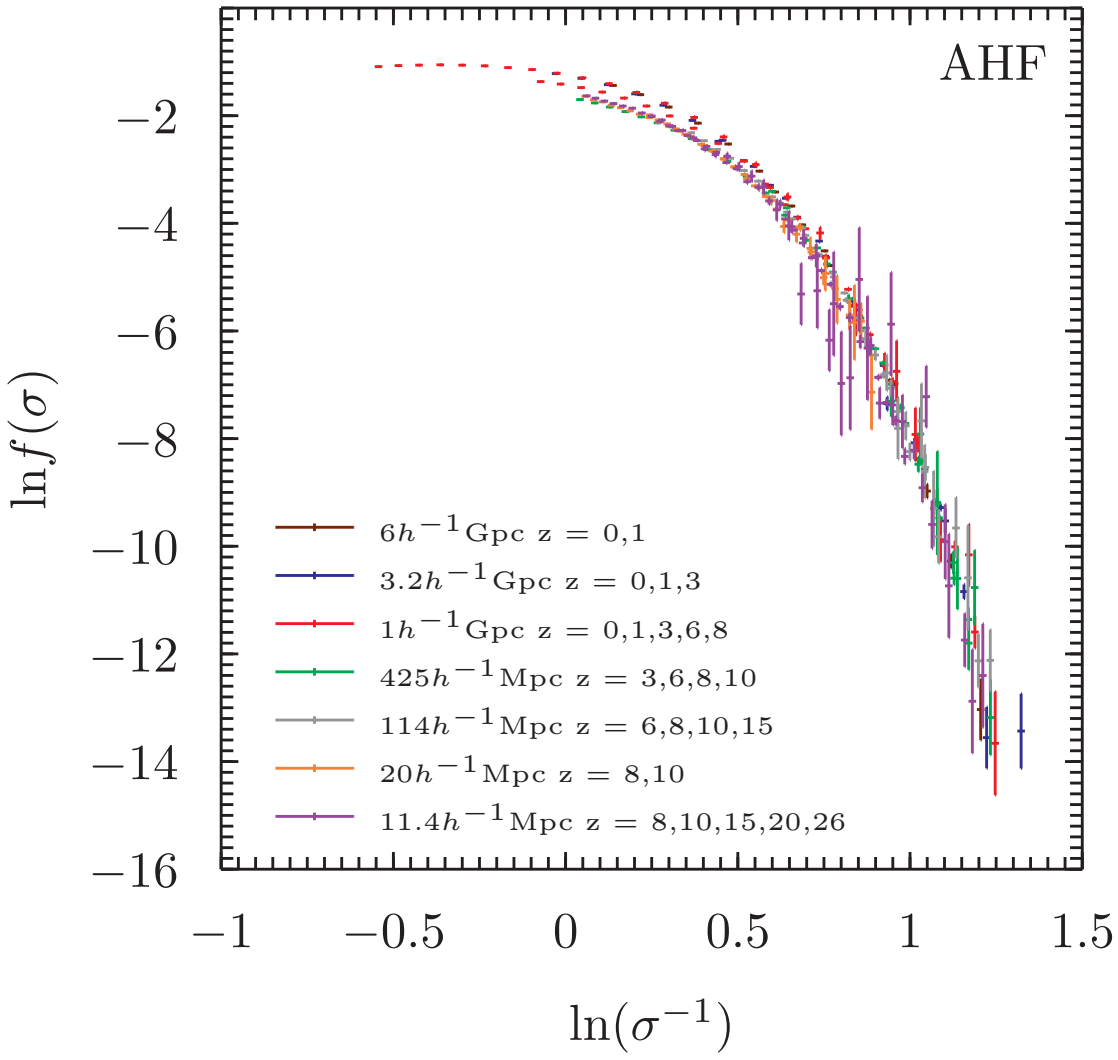


Figure 3.4: The AHF mass function across all simulations and redshifts.

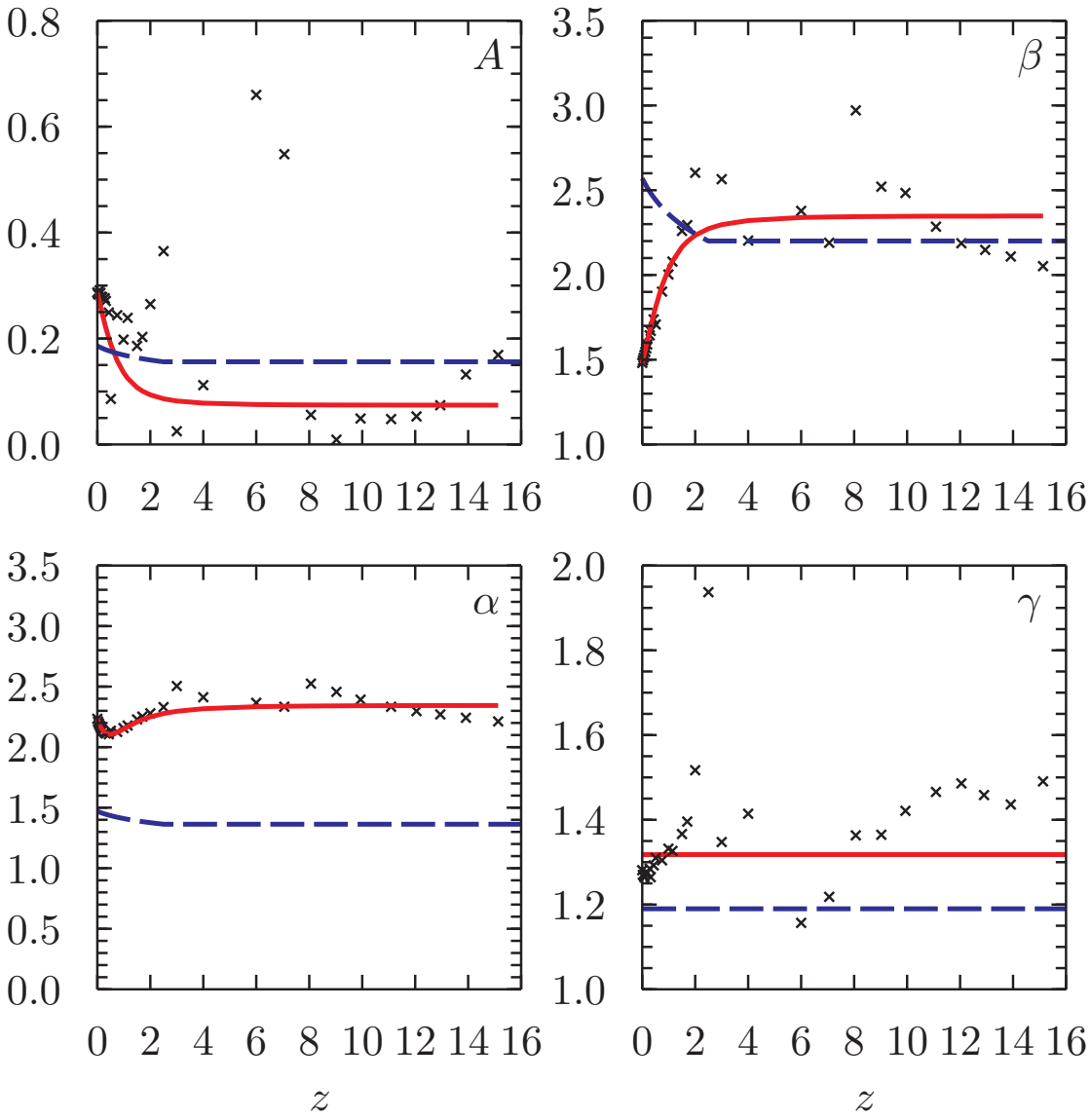


Figure 3.5: Evolution of the parameters in equation 3.14 for CPMSO haloes. The Tinker et al. (2008) values are shown in blue for comparison, the model used in this work is shown in red. The parameters have been fitted in the order: $\gamma \rightarrow A \rightarrow \alpha \rightarrow \beta$.

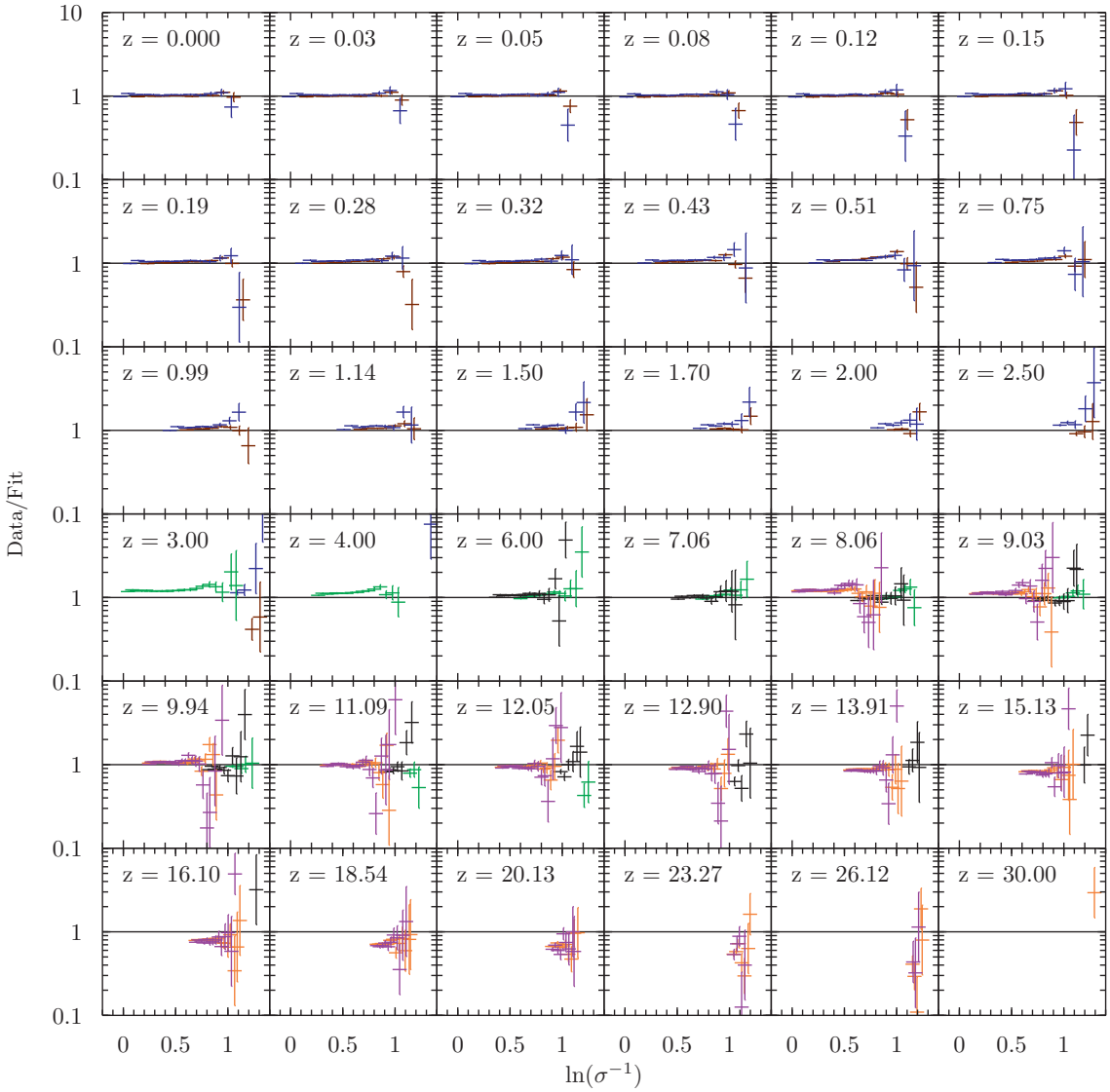


Figure 3.6: Ratios between the CPMSO redshift-dependent mass function and the data from the CPMSO halofinder across 36 redshifts from $z = 0 - 30$. The colour scheme is as per Figure 3.3.

We perform a least-squares fit on the CPMSO data for 36 output redshifts, initially allowing all of the parameters in equation 3.14 to vary. Tinker et al. (2008) previously found, based on lower-redshift data ($0 < z < 2$), that the parameter controlling the exponential cut-off scale, γ, is approximately constant across their range of redshifts. We find a similar trend albeit at a slightly different value of γ from Tinker et al. (2008) (see Figure 3.5, bottom-right panel). Given this result we proceed by fixing this parameter at its value in the approximation of universality: i.e. $\gamma = 1.318$. We then fit the CPMSO data across all redshifts again. We find that using a parameterisation that includes $\Omega_m(z)$, the matter content of the universe at a given redshift, enables us to capture late time behaviour that differs from that at high-redshift. This gives us a fit for $A(z)$ in the following form:

$$A(z) = \Omega_m(z) \left\{0.990 \times (1+z)^{-3.216} + 0.074\right\}. \tag{3.15}$$

With γ and A modelled, we repeat the procedure for α and then β. We find for $\alpha(z)$:

$$\alpha(z) = \Omega_m(z) \left\{5.907 \times (1+z)^{-3.599} + 2.344\right\}, \tag{3.16}$$

and for $\beta(z)$:

$$\beta(z) = \Omega_m(z) \left\{3.136 \times (1+z)^{-3.058} + 2.349\right\}. \tag{3.17}$$

We show how our model compares to the parameter fitting data in Figure 3.5.

In Figure 3.6 we compare our redshift-dependent fit to our CPMSO data across all redshifts from $z = 0$ and $z = 30$. We find that the fit is excellent at low redshifts and remains quite good, within 20%, all the way to $z \sim 20$. For $z > 20$ our fit still gives a reasonable match but the data have large error bars due to the scarcity of haloes then. Nonetheless, we note that around $z = 3 - 4$ our fit slightly under-predicts the abundances of SO haloes across the lower-mass $\ln\sigma^{-1}$ range we cover. There is also an apparent under-prediction for lower $\ln\sigma^{-1}$ around $z = 8$ and there is perhaps an over-prediction at very high redshifts ($z > 15$).

3.3.5.2 AHF-based fits

Our CPMSO on-the-fly halo finder is by its nature simplified and potentially more approximate than AHF. In Figure 3.7 we show the ratio between the mass functions derived

using these two SO-based halofinders, AHF and CPMSO. In Figure 3.8 we show the ratio between the FOF mass function and the AHF mass function. Both AHF and FOF exhibit consistent behaviour across all simulation volumes and resolutions and for all redshifts. The CPMSO halo finder largely agrees with AHF, typically within 10-20%, and much less at low redshifts. There are some systematic differences around $z = 8$ in the transition between the 20 h^{-1}Mpc and 114 h^{-1}Mpc boxes. On the other hand, while both FOF and AHF are consistent across all box sizes, there is a systematic trend for the FOF to yield more rare, massive haloes than AHF for all volumes and redshifts. A similar trend was noted previously by Reed et al. (2007) and Tinker et al. (2008). Based on these results, we conclude that the AHF (and FOF) data are more consistent across different box sizes than the CPMSO halofinder.

Given these differences between the two SO halo finders, in addition to the CPMSO fit above, we also provide AHF fits to the data, with three different types: 1) a redshift-dependent one, based around the redshift evolution we have observed in the CPMSO mass function, but re-normalised to the AHF data; 2) a present-day one, based on our results at $z = 0$, for direct comparison with previous works and 3) a more precise high-redshift, "Epoch of Reionization (EoR)" halo mass function, based on all the AHF data past $z = 6$.

In light of the relationship between the two SO halofinders in Figure 3.7, we expect that the redshift parameterisation presented in equations 3.15–3.17 applied to AHF data leads to a mass function that is roughly correct. The fit to AHF at $z = 0$ can be improved by introducing a slightly different parameterisation for A, as follows:

$$A(z) = \Omega_m(z) \left\{1.097 \times (1+z)^{-3.216} + 0.074\right\}. \tag{3.18}$$

The combination of equations 3.16, 3.17, 3.18, and $\gamma = 1.318$ results in the ratios shown in Figure 3.9. We see that this fit is accurate to $\sim 10\%$ for redshifts less than $z = 15$. The shape of the function in the low $\ln\sigma^{-1}$ range is slightly different for $z > 8$, although the amplitude and the shape are both correct for $\ln\sigma^{-1} > 0.3$ at $z = 8 - 10$.

We now provide more accurate parameterisations for our AHF data for several different redshift ranges. At $z = 0$ we have: $A = 0.194$, $\alpha = 2.267$, $\beta = 1.805$ and $\gamma = 1.287$ valid in the range $-0.55 \leq \ln\sigma^{-1} < 1.05$. As the redshift evolution of the spherical overdensity mass function is mainly apparent at later times ($z < 3$), we can provide a universal fit that is appropriate for high redshift studies, for example for probing the EoR and the Cosmic Dark Ages. An AHF mass function fitted to all our data from $z = 6$ upwards has the

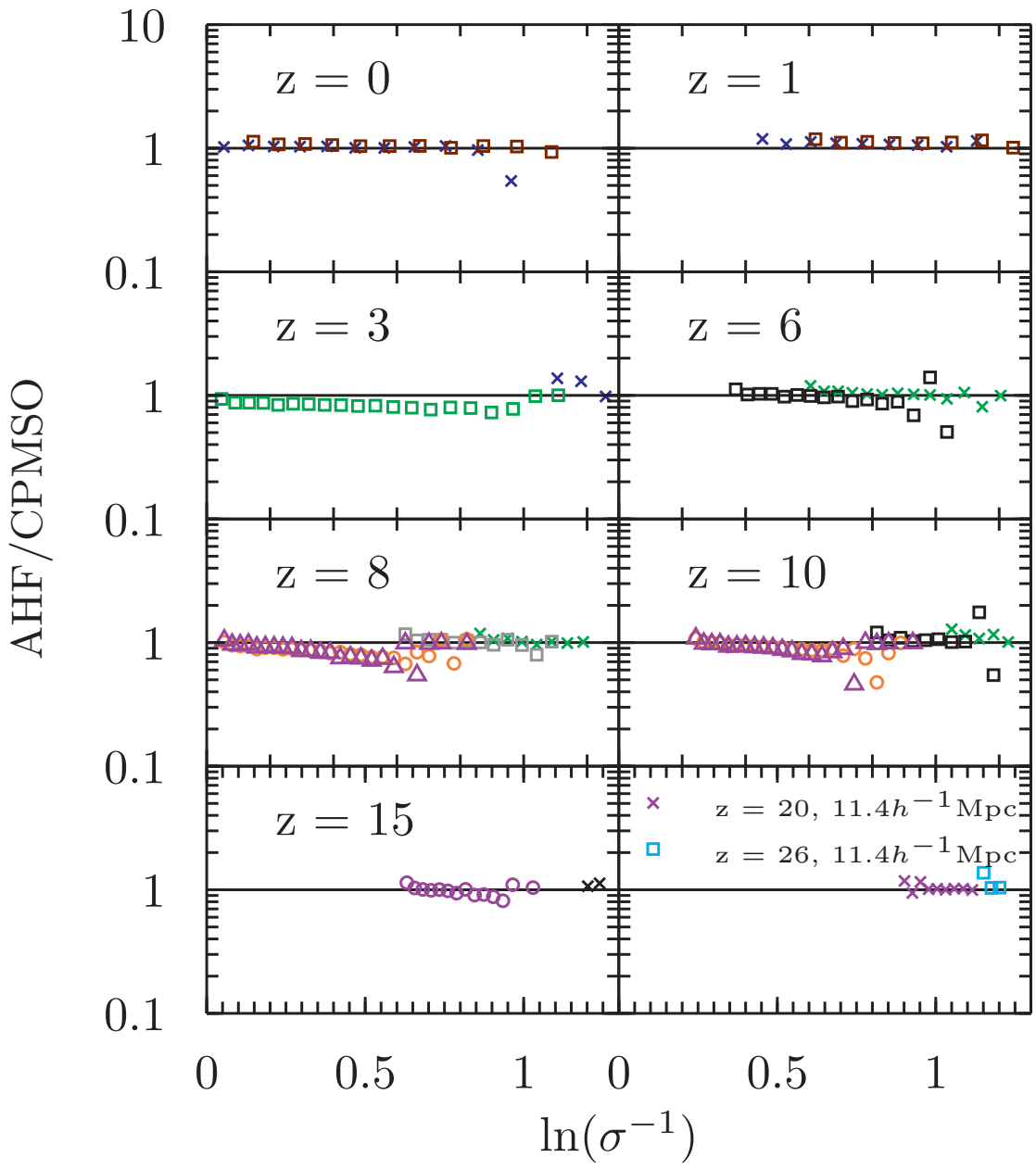


Figure 3.7: Comparison of the mass functions from the CPMSO and AHF halofinders across redshifts from $z = 0 - 26$. The colour scheme matches the simulations as per Figure 3.3

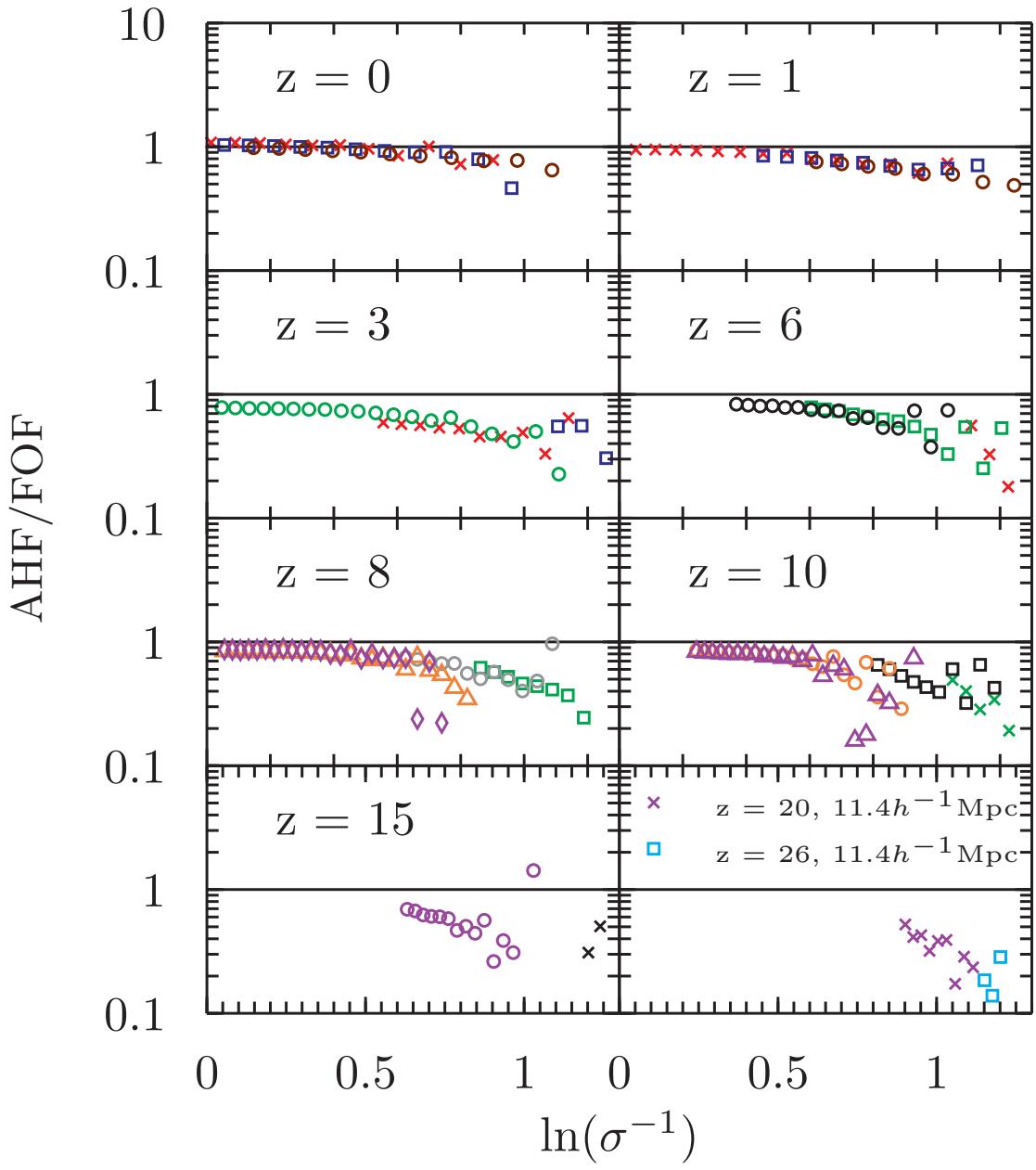


Figure 3.8: Comparison of the mass functions from the FOF and AHF halofinders across redshifts from $z = 0 - 26$. The colour scheme matches the simulations as per Figure 3.3

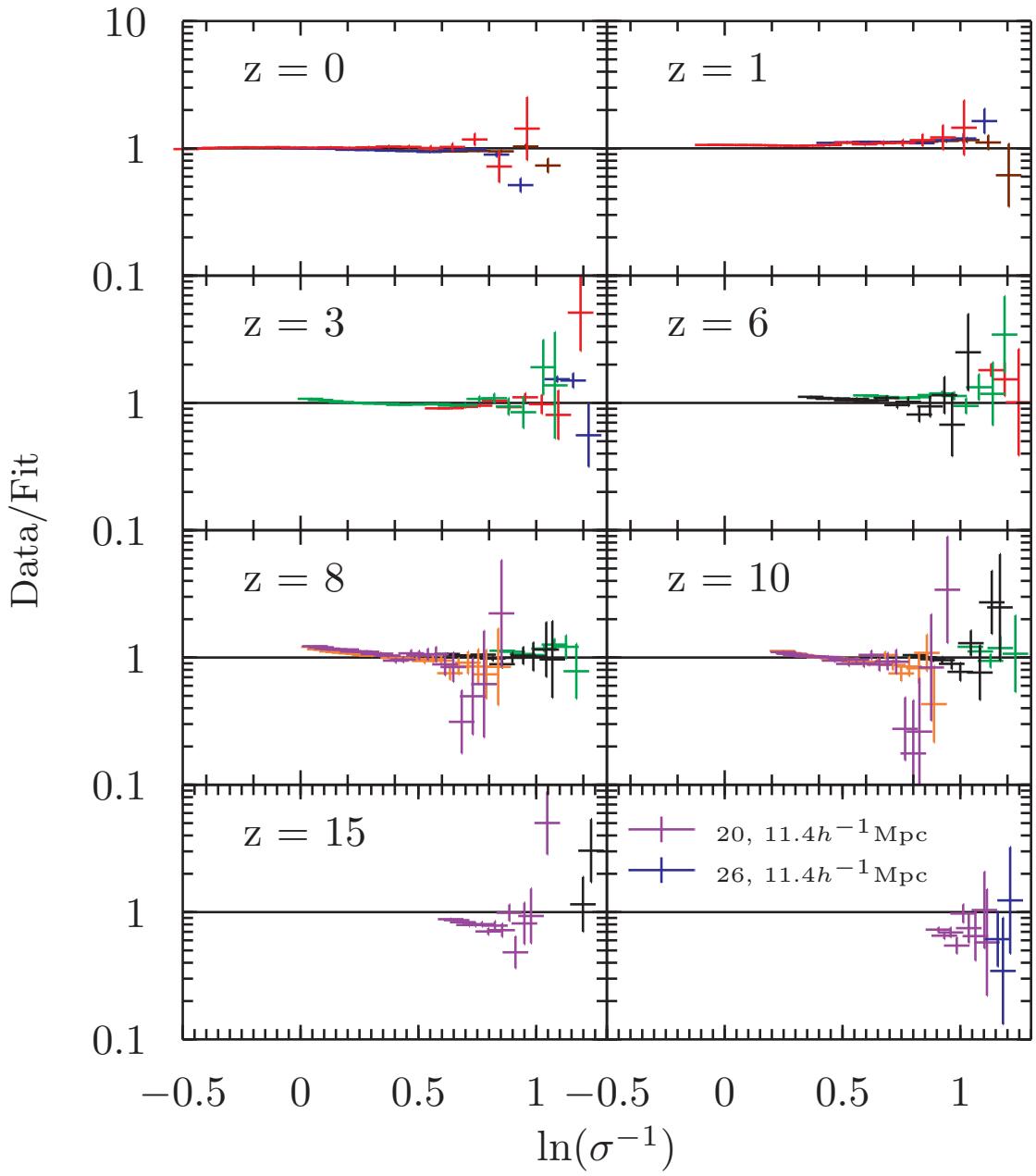


Figure 3.9: Ratios between the amplitude-modified CPMSO mass function and AHF data from $z = 0$ to $z = 26$. The colour scheme is as per Figure 3.3.

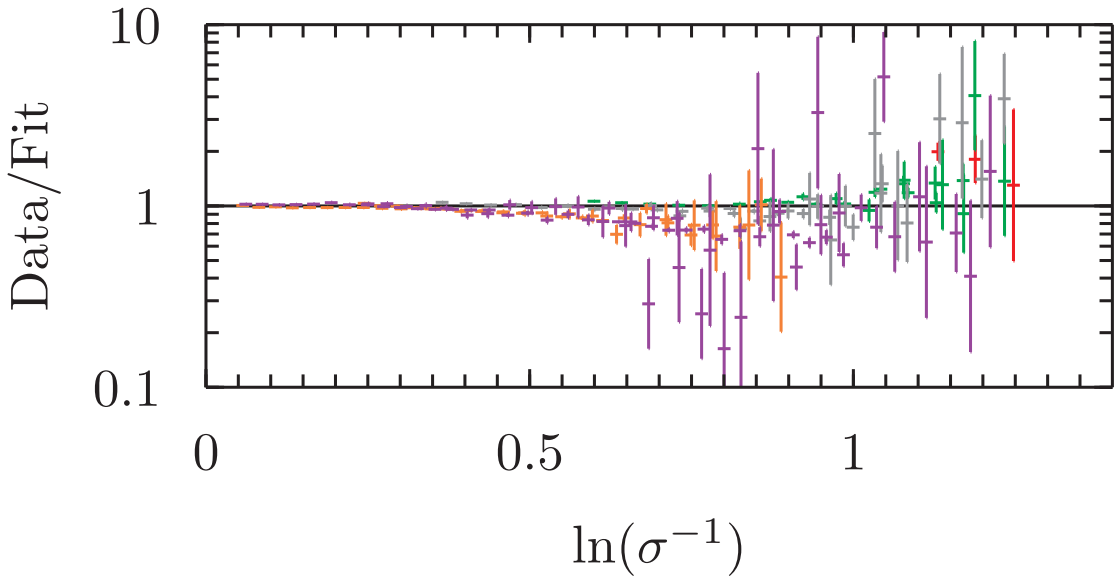


Figure 3.10: Ratio between the EoR AHF fit and simulation data from $z = 6 - 26$. The colour scheme is as per Figure 3.3.

following parameterisation: $A = 0.563$, $\alpha = 3.810$, $\beta = 0.874$ and $\gamma = 1.453$ valid in the range $-0.06 \leq \ln\sigma^{-1} < 1.24$ (corresponding at redshift $z = 8$ to a mass range of 3.5×10^6 – 6.3×10^{11} $h^{-1}\mathrm{M}_\odot$). These values are notably different to those given by our $z = 0$ redshift parameterisation because of the lack of constraining data for low values of $\ln\sigma^{-1}$. The ratio of the EoR fit versus our data is shown in Figure 3.10. There is considerable scatter in the high $\ln\sigma^{-1}$ end due to data being incorporated from high redshifts when haloes are scarce. For the range $0.05 \leq \ln\sigma^{-1} < 0.5$ the fit is accurate to within 10%.

3.3.6 Varying Δ

For practical purposes it is useful to have a simple method for adjusting the spherical overdensity mass function fits derived above to account for haloes defined with different overdensities. To this end we have run AHF on the 1 h^{-1}Gpc box using values for Δ ranging between 100 to 1600, at $z = 0$, 1 and 3. The data have then been fitted with Δ incorporated into the parameterisation. This allows, to good precision at lower redshifts, a mass function based on a given Δ to be inferred from our Δ_{178} fits. The results from this procedure are shown in Figure 3.11.

The parameterisation is as follows. We assume that the $f_{\Delta=178}$ mass function is suitable for describing mass functions with different choices for Δ when it is suitably scaled by a function, $\Gamma(\Delta, \sigma, z)$:

$$f_\Delta = \Gamma(\Delta, \sigma, z) f_{\Delta=178}. \tag{3.19}$$

We find that the following form for Γ is suitable for describing our data:

$$\Gamma(\Delta, \sigma, z) = C(\Delta) \left(\tfrac{\Delta}{178}\right)^{d(z)} \exp\left[p \left(1 - \tfrac{\Delta}{178}\right) / \sigma^q\right], \tag{3.20}$$

where

$$\begin{aligned} C(\Delta) &= \exp\left[0.023 \left(\tfrac{\Delta}{178} - 1\right)\right], \\ d(z) &= -0.456 \Omega_m(z) - 0.139, \\ p &= 0.072, \\ q &= 2.130. \end{aligned} \tag{3.21}$$

The redshift-dependence of the fit is based solely on $\Omega_m(z)$, via the $d(z)$ parameter.

3.4 Summary and discussion

3.4.1 Fitted functions

We have provided a number of fitted mass functions based on our data. These are summarised in Table 3.1.

3.4.2 Comparison to existing fits

In Figure 3.12 we show our universal FOF mass function relative to a number of fits available in the literature at $z = 0$, including the widely-used Press-Schechter (PS) (Press & Schechter, 1974), Sheth-Tormen (ST) (Sheth & Tormen, 2002) and Jenkins et al. (Jenkins et al., 2001) fits, as well as a number of more recent fits (Warren et al., 2006; Reed et al., 2003, 2007; Crocce et al., 2010; Courtin et al., 2011; Bhattacharya et al., 2011; Angulo et al., 2012). To produce this plot we have assumed that the underlying power spectra – i.e. the underlying cosmologies – of each of the fits is the same as our cosmology A (3.3). This assumption is necessary to compare the fits to $z = 0$ halo masses, and we stress that there would exist small departures from this plot in the precise case if the actual cosmologies of all the independent fits were employed. We note that the fits by Reed et al. (2007),

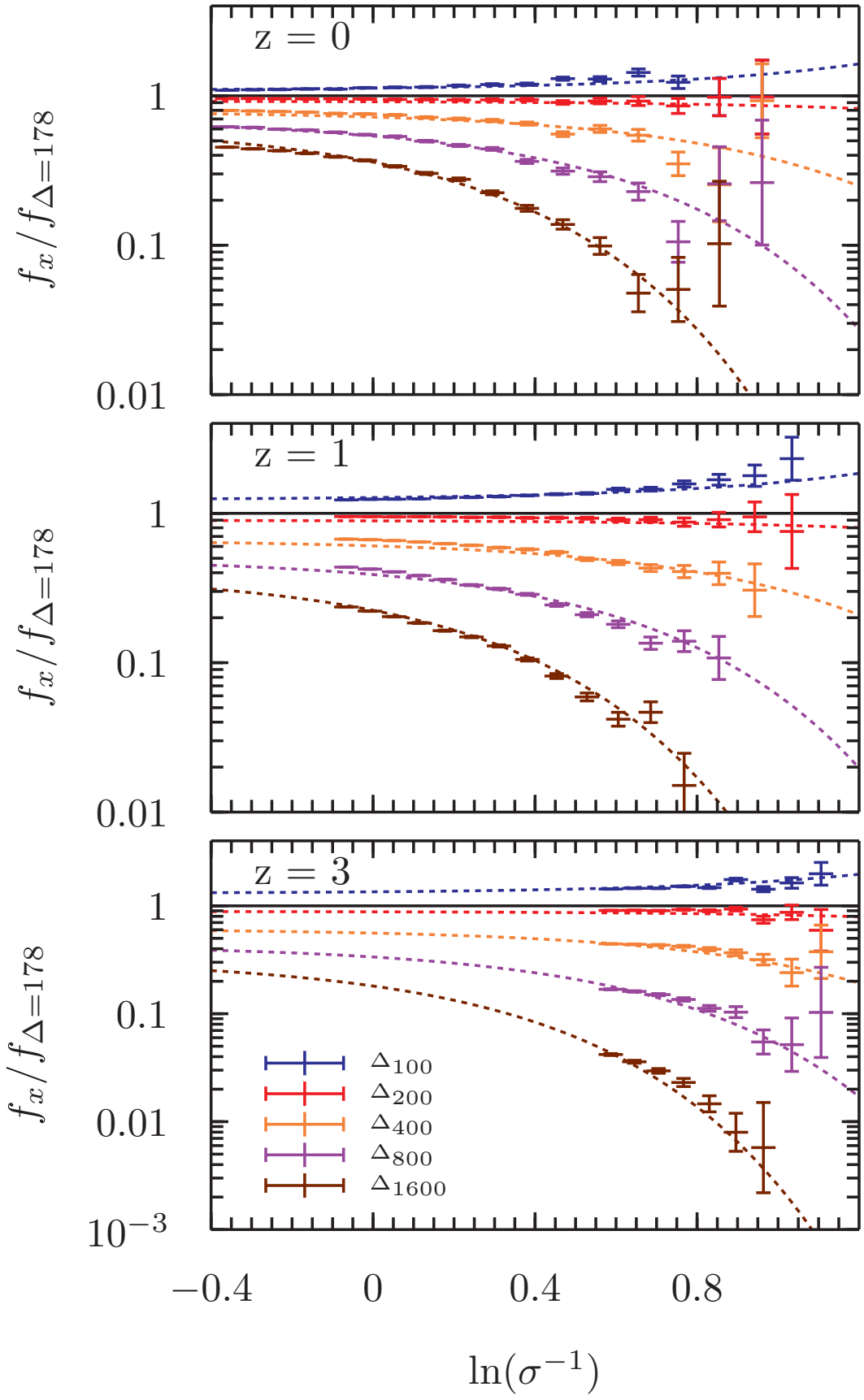


Figure 3.11: The effect of the choice of the overdensity criterion, Δ, on the mass function. For z = 0,1,3 (top panel to bottom respectively), the ratio of mass functions with a variety of choices for Δ are shown relative to the Δ_{178} function. All data have been calculated from the 1 h^{-1}Gpc simulation. Fitted curves, described in the text, are shown as dashed lines.

Table 3.1: Mass function fitting parameters. For details on the redshift evolution of these parameters see sections 3.3.5.2

	FOF *Uni.* $z = all$	CPMSO $z = 0$	CPMSO +AHF $z = 0$	AHF $z = 0$	AHF *EoR* $z = 6+$
A	0.282	0.287	0.316	0.194	0.563
α	2.163	2.234	2.234	1.805	3.810
β	1.406	1.478	1.478	2.267	0.874
γ	1.210	1.318	1.318	1.287	1.453
z dep.	No	Yes	Yes	No	No

Crocce et al. (2010) and Bhattacharya et al. (2011) are redshift-dependent, while Press & Schechter (1974); Sheth & Tormen (2002) and Courtin et al. (2011) are parameterised using $\delta_c(z)$. Aside from the classic, and less precise Press & Schechter (1974) expression, our fit agrees to $\sim 10\%$ with the others for most of the mass range and, for very massive clusters ($> 10^{15}\ h^{-1}\mathrm{M}_{\odot}$), it matches very closely the recent Angulo et al. (2012) fit based on the Millennium series of simulations. The effect of using data across all redshifts to create the universal fit is readily seen as the fit versus the $z = 0$ data (shown in Figure 3.14) is seen to be under-predicting mid-sized haloes. In fact, our $z = 0$ data match well a number of the predictions from the literature for these haloes. At high masses our data (Figure 3.14) illustrate the large scatter from shot noise that is typically observed in the tail of the mass function.

With the exception of the PS, ST and Reed et al. (2003) results, the largest discrepancy both with respect to our fit and overall scatter is for very large, rare haloes, whose statistics in most simulations is poor. A reliable result in this range requires very large simulation volumes of tens to hundreds of Gpc^3, comparable to the volume of the observable universe.

As we discussed above, the only other robust SO-based halo mass function fit is the one provided by Tinker et al. (2008). In Figure 3.13 we show the ratio between our AHF $z = 0$ fit and the Tinker et al. (2008) fit at $z = 0$. We also show how the CPMSO, redshift-dependent mass function compares to the AHF $z = 0$ fit and the ratio between the AHF $z = 0$ fit and the $\Delta = 200$ adjusted AHF $z = 0$ fit (calculated using equation 3.19). The

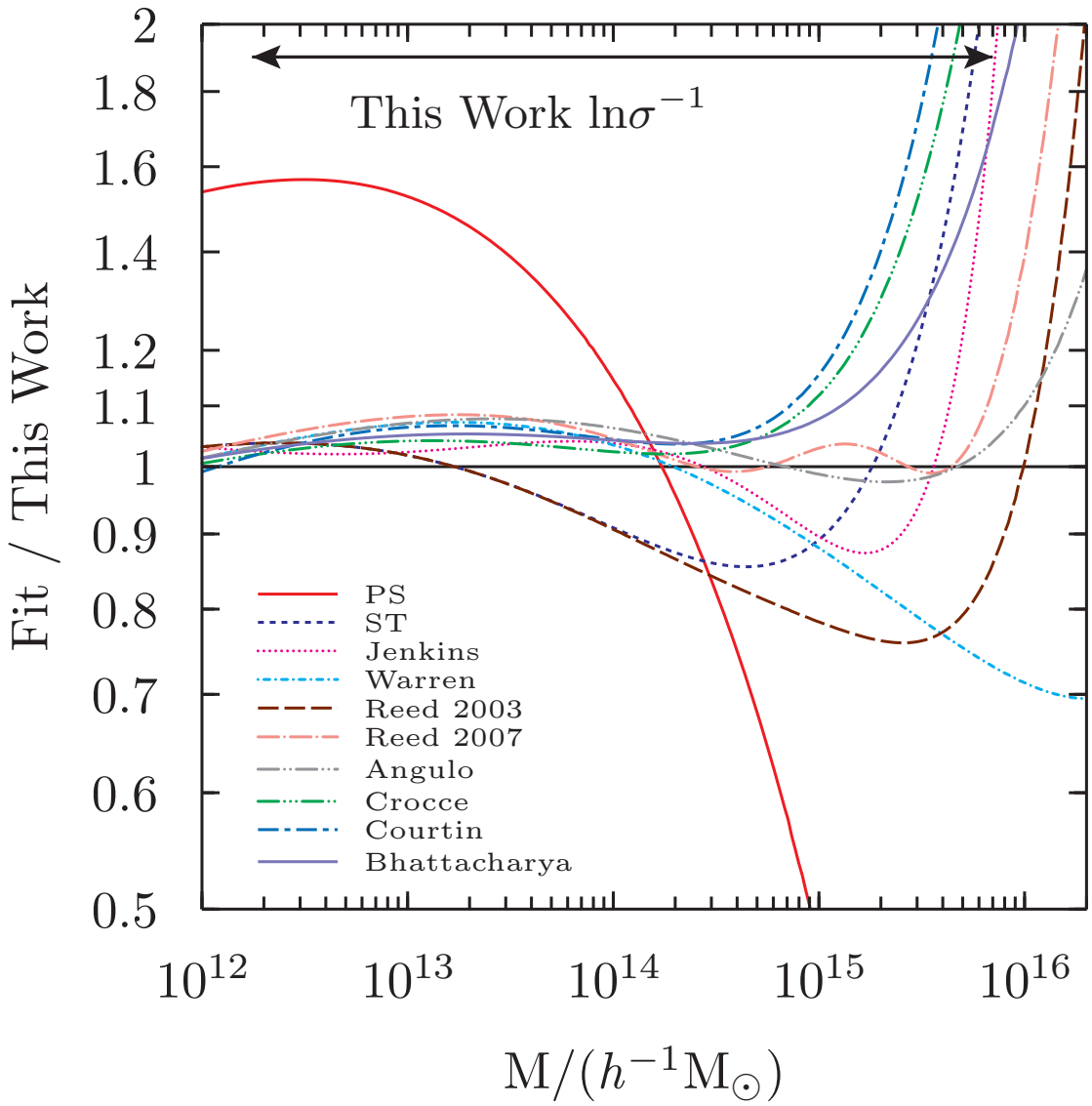


Figure 3.12: Ratios between the universal FOF fit and a number of FOF-based fits from the literature for $z = 0$. We have assumed an underly cosmology identical to cosmology A from § 3.3.

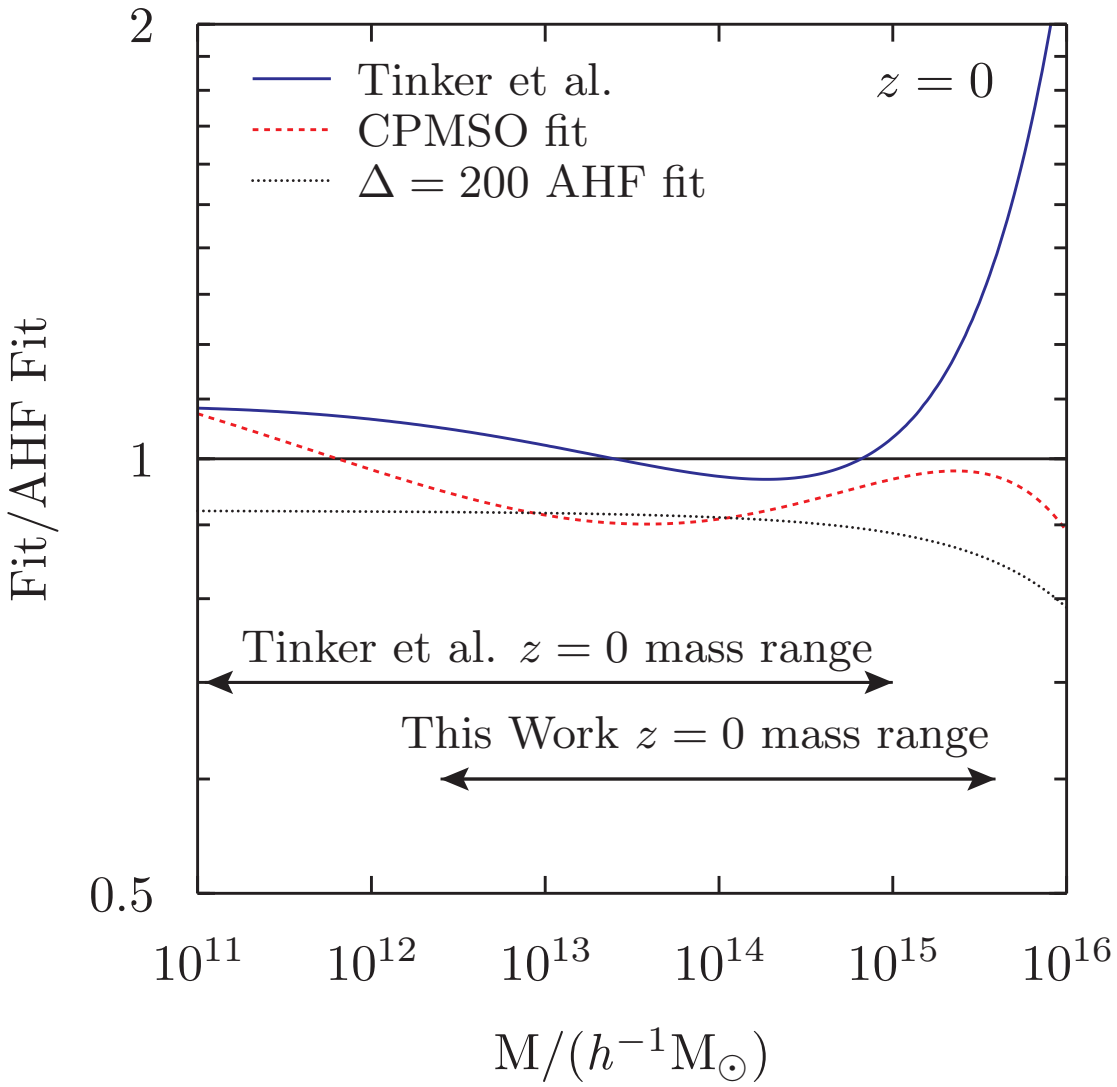


Figure 3.13: Ratios, at $z = 0$, of the AHF $z = 0$ fit (solid black line) versus the Tinker et al. (2008) fit (solid blue line), the redshift parameterised CPMSO fit (dotted red line) and the AHF precise fit adjusted to $\Delta = 200$ using equation 3.19 (dotted black line).

Tinker et al. (2008) fit is based on data in the range $-1.4 < \ln\sigma^{-1} \leq 0.9$ for $z < 0.2$, while our fit includes data over the range $-0.55 < \ln\sigma^{-1} \leq 1.35$ across all redshifts or specifically $-0.55 < \ln\sigma^{-1} \leq 1.05$ at $z = 0$.

We show an agreement with Tinker et al. (2008) to $\sim 5\%$ for haloes in the mass range $10^{12} - 10^{15}\ h^{-1}\mathrm{M}_{\odot}$. Past $\sim 10^{15}\ h^{-1}\mathrm{M}_{\odot}$ we predict a lower collapsed mass fraction, although this is outside the range of the Tinker et al. (2008) fitting data. The Tinker et al. (2008) fit is based on Δ_{200} haloes and we show how our Δ_{200} prediction via equation 3.19 compares to the underlying precision AHF plot. We note that this is slightly less congruent with the Tinker et al. (2008) fit than the Δ_{178} prediction, although it is still in agreement to within 10% for most of the mass range in question.

3.4.3 Universality

The results presented here are in line with other recent studies of the halo mass function that have addressed the question of universality, i.e. whether the mass function can be considered invariant over all redshifts or cosmological models when expressed in a suitable form. It is beyond the scope of this chapter to discuss invariance under different cosmologies, for work on this see Jenkins et al. (2001); Warren et al. (2006); Tinker et al. (2008) and in particular Courtin et al. (2011) and Bhattacharya et al. (2011). The mass function has been found to depend only weakly on cosmology when couched in terms of FOF haloes with a fixed linking length. More study is required to address the question of how SO halo mass functions differ in varying cosmologies.

We find that the assumption of halo mass function universality across redshift is approximately valid but is violated on closer inspection, in qualitative agreement with the conclusions of other studies. White (2002) highlighted that non-universality existed for most halo mass estimators; Tinker et al. (2008) present an SO mass function that includes redshift-dependent parameters for $z \leq 2.5$; similarly Crocce et al. (2010) and Bhattacharya et al. (2011) present FOF mass functions with redshift-dependent parameters for $z \leq 2$; Reed et al. (2007) produce a high-redshift mass function that contains a redshift-dependence via the effective slope of the power spectrum at the scale of the halo radius. We observe here that the FOF halo mass function for fixed linking length is close to universal for a very large redshift range (from $z = 26$ to the present). Jenkins et al. (2001) noted that taking a constant linking length keeps the mass function closer to a universal form, a result corroborated by Lukic et al. (2007), who found that even for high redshifts the mass function of Warren et al. (2006) was suitable for FOF haloes, despite it being

calibrated on $z = 0$ data. While we derived a universal fit for the FOF halo mass function for fixed linking length of 0.2, there clearly is some, albeit modest, redshift evolution about this fit. We show in Figure 3.14 the residuals between our FOF universal fit and our data for nine different redshifts. The amplitude of the data relative to the fit drops away markedly for lower $\ln\sigma^{-1}$ values, to the extent that by $z \sim 8$ and above the data are around 20% lower than the prediction around $\ln\sigma^{-1} \sim 0.5$.

In contrast, the SO-based halo mass functions are clearly not universal over the entire redshift range we consider, which was also previously noted by Tinker et al. (2008), based on data over a slightly larger range in and σ than our data, albeit only for $z < 2$. Here, we presented SO halo mass function fits that are applicable across a large range of redshifts, by explicitly introducing an Ω_m- and redshift-dependent parameterisation of the mass function.

3.4.4 Halofinder comparison

3.4.4.1 Relationship between SO and FOF haloes

The differences between an FOF halo and an SO halo, and the respective mass functions derived from haloes found using the two algorithms, are significant. A typical FOF halo is arbitrarily shaped and has been described in the literature as a demarkation of an isodensity contour in real space, with a value of Δ_{FOF} – the overdensity of matter at the halo boundary – that is under some debate. Estimates for Δ_{FOF} include $\Delta_{FOF} \approx 2b^{-3}$, i.e $\Delta_{FOF} \approx 250$ for the standard choice of $b = 0.2$, (Navarro et al., 1997); $\Delta_{FOF} \approx 0.48b^{-3}$, i.e. $\Delta_{FOF} \approx 60$ (Lacey & Cole, 1994; Summers et al., 1995; Audit et al., 1998); $\Delta_{FOF} \approx 74$ (Warren et al., 2006); and $\Delta_{FOF} \approx 81.62$ More et al. (2011), all for $b = 0.2$. In contrast the SO algorithm produces haloes that are by definition spherical and that adhere to a strict overdensity criterion, e.g. $\Delta = 178$ times the background matter density. This overdensity refers to the average density of matter contained within the halo. Haloes are generally very centrally-concentrated, thus the overdensity of matter at the boundary of an SO halo is much lower than the mean overdensity and the relation between the two is dependent on the profile of the halo in question. More et al. (2011) note that for a singular isothermal sphere (SIS) density profile, $\rho(r) \propto r^{-2}$, an overdensity at a halo's boundary of $\Delta_{edge} \sim 60$ corresponds to an enclosed overdensity of $\Delta_{halo} \sim 180$. This loosely links the value of b in the FOF algorithm to Δ in the SO algorithm –by assuming that the FOF halo is spherical– but the simplified SIS profile does not reproduce simulated dark matter haloes well. More appropriate profiles include the TIS profile (Shapiro et al., 1999; Iliev & Shapiro, 2001),

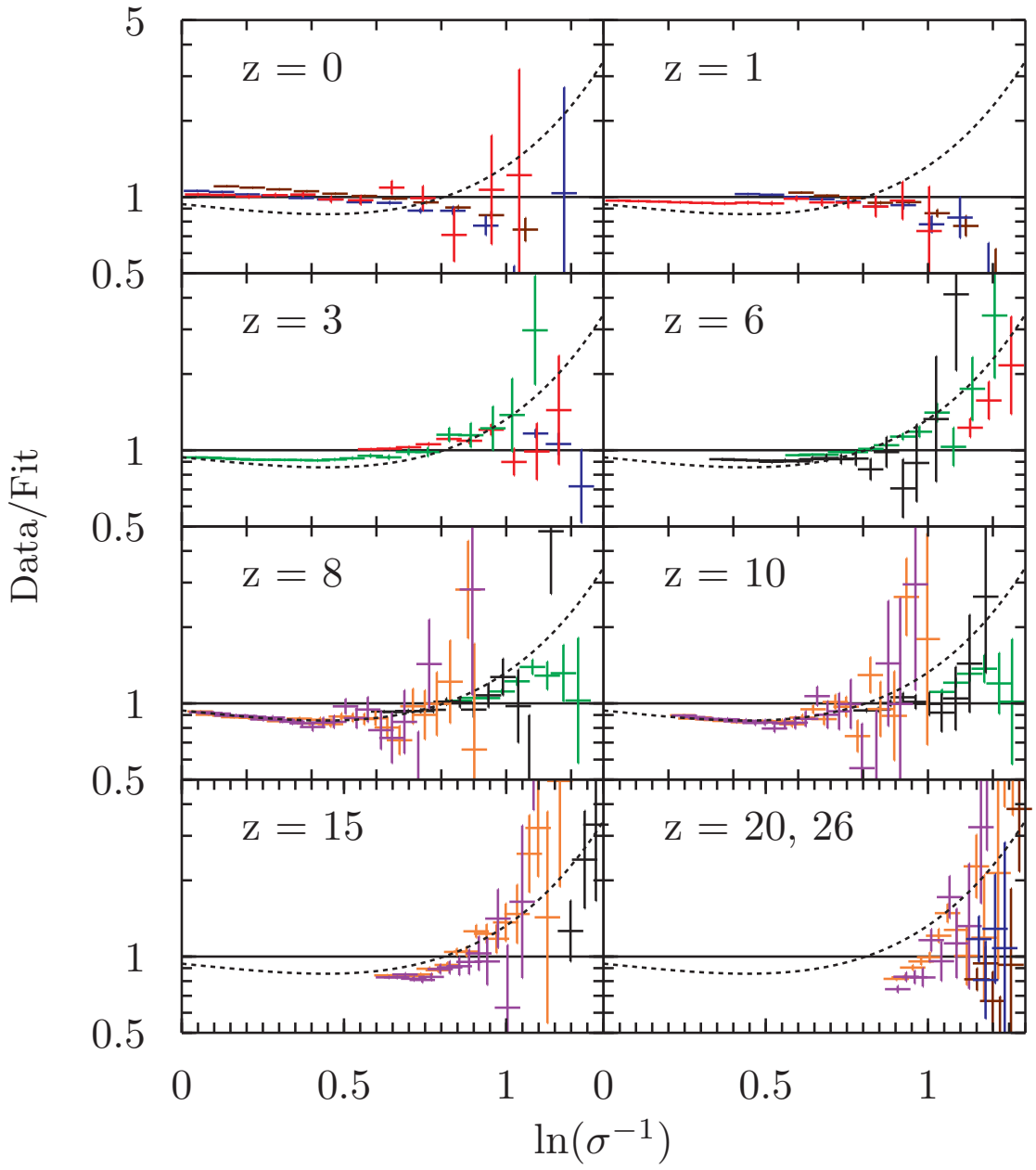


Figure 3.14: Redshift evolution of the FOF mass function versus the universal fit. The fit of Sheth & Tormen (2002) is shown as a dashed line for reference.

the Einasto profile (Navarro et al., 2004; Gao et al., 2008; Navarro et al., 2010) and the NFW profile. More et al. (2011) used the latter, sampling NFW haloes with random particle realisations, and show that their value of $\Delta_{FOF} \approx 81.62$ corresponds to a range of Δ_{halo} between $\approx 200-600$, depending on the concentration parameter of the NFW haloes. Lukic et al. (2009) use a similar approach and were able to recover, to 5% accuracy, an SO mass function from FOF haloes by individually relating the FOF halo masses to their SO counterparts. The translation between the two was based on the concentration parameter of the haloes' NFW profiles and three empirically fitted parameters.

3.4.4.2 Choice of linking length and overdensity criterion

Given these analyses, it is not surprising to observe a marked difference in the physical structure of FOF and SO haloes. Figure 3.15 shows a large halo from our 20 h^{-1}Mpc simulation at $z = 8$ captured by the CPMSO and the FOF halo finders, respectively. We show the FOF halo found with linking lengths of 0.2, 0.15, and 0.1 in grey, blue and red respectively. This is the second largest halo in our volume and it illustrates particularly well the differences between the two algorithms. For this halo at this redshift a linking length of 0.2 is far too aggressive and significant overlinking has occurred.

In Figure 3.16 we compare the halo mass functions based on FOF haloes with linking lengths of 0.2, 0.15, 0.1 and SO haloes with overdensity choices of 100, 178, 200 and 800, at $z = 0$ from the 1 h^{-1}Gpc box (top panel) and at $z = 8$ from the 114 h^{-1}Mpc box (bottom panel). The $z = 0$ result shows that, for a suitable range of σ, an overdensity of 178 is comparable to a linking length of 0.2 to within 10%. However, for a 0.2 linking length there appears to be a trend towards lower overdensities for the higher mass haloes. This is likely due to the increasing influence of overlinking on the masses of the larger FOF haloes. The $z = 8$ results show that a linking length of 0.2 is not at all complementary to an overdensity choice of 178 in the EdS regime of structure growth. In fact, we see that an overdensity of 100 is more consistent with a linking length of 0.2 at $z = 8$. The equivalent linking length for an overdensity of 178 lies between 0.2 and 0.15 and much closer to 0.15 than to 0.2 and a linking length of 0.1 roughly corresponds to overdensity of 800. These results are consistent with the result of Cohn & White (2008), namely that, at $z = 10$, the mass functions of FOF and SO haloes are similar when an overdensity choice of $\Delta = 180$ and a linking length of $ll = 0.168$ are used. Given these results, we expect there to be an evolution in the relationship between overdensity and linking length. The empirical results of Lukic et al. (2009), Courtin et al. (2011) and More et al. (2011) all contain a

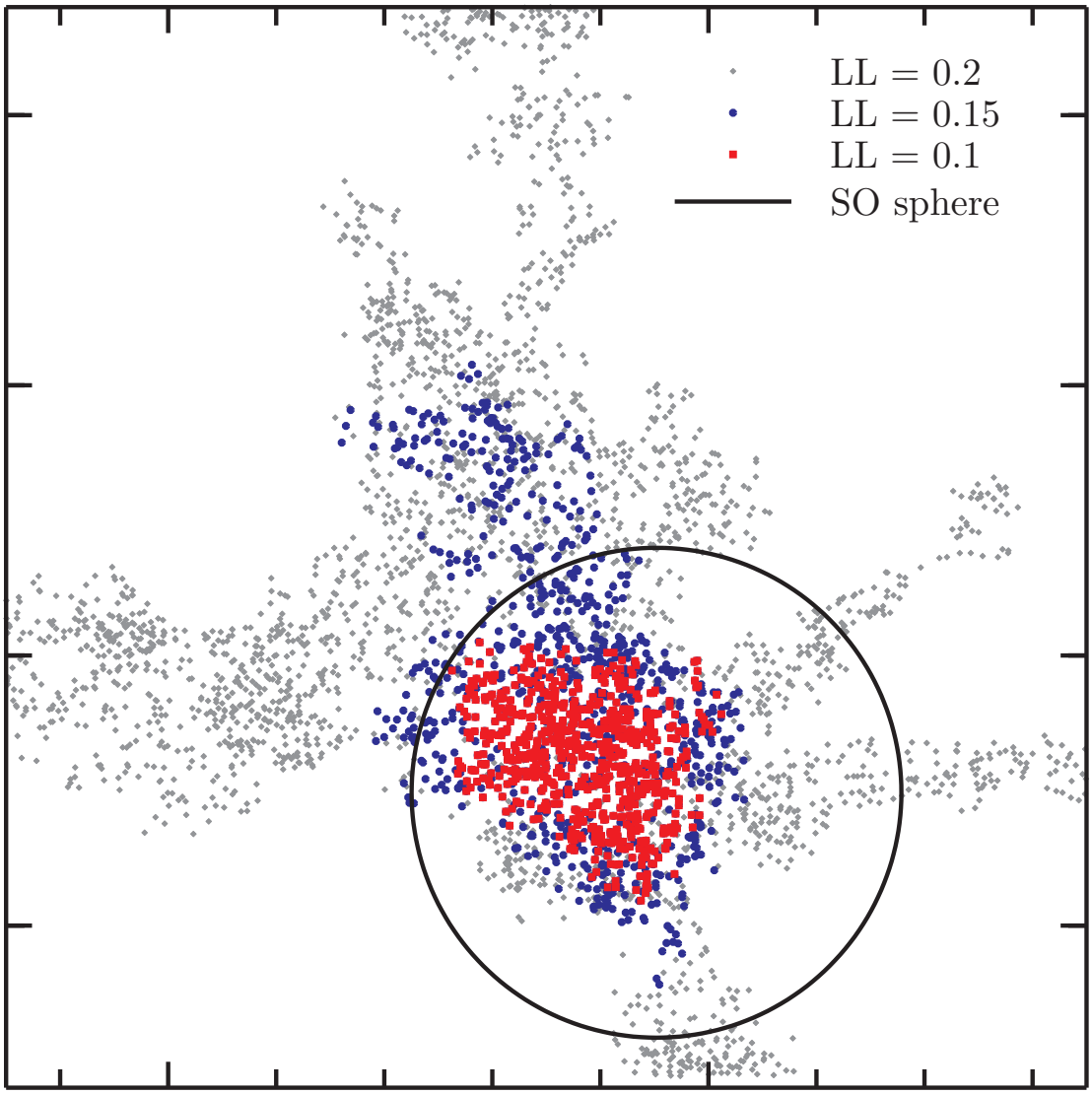


Figure 3.15: Image of a large halo in the 20 h^{-1}Mpc box at $z = 8$. The circle represents the extent of the Δ_{178} cutoff used in CPMSO. The z direction has been projected onto the x-y plane. The CPMSO halo mass is $3.1{\times}10^{10}$ $h^{-1}\mathrm{M}_{\odot}$ and it contains 9.3 million particles. The dots represent aggregations of at least 20 particles found in the FOF version of the same halo. Grey shows the halo captured with a linking length of 0.2, blue 0.15 and red 0.1. The masses (particle counts) are $4.8{\times}10^{10}$ $h^{-1}\mathrm{M}_{\odot}$ (13.1 million), $3.7{\times}10^{10}$ $h^{-1}\mathrm{M}_{\odot}$ (8.9 million) and $2.1{\times}10^{10}$ $h^{-1}\mathrm{M}_{\odot}$ (5.8 million) for $b = 0.2, 0.15, 0.1$ respectively.

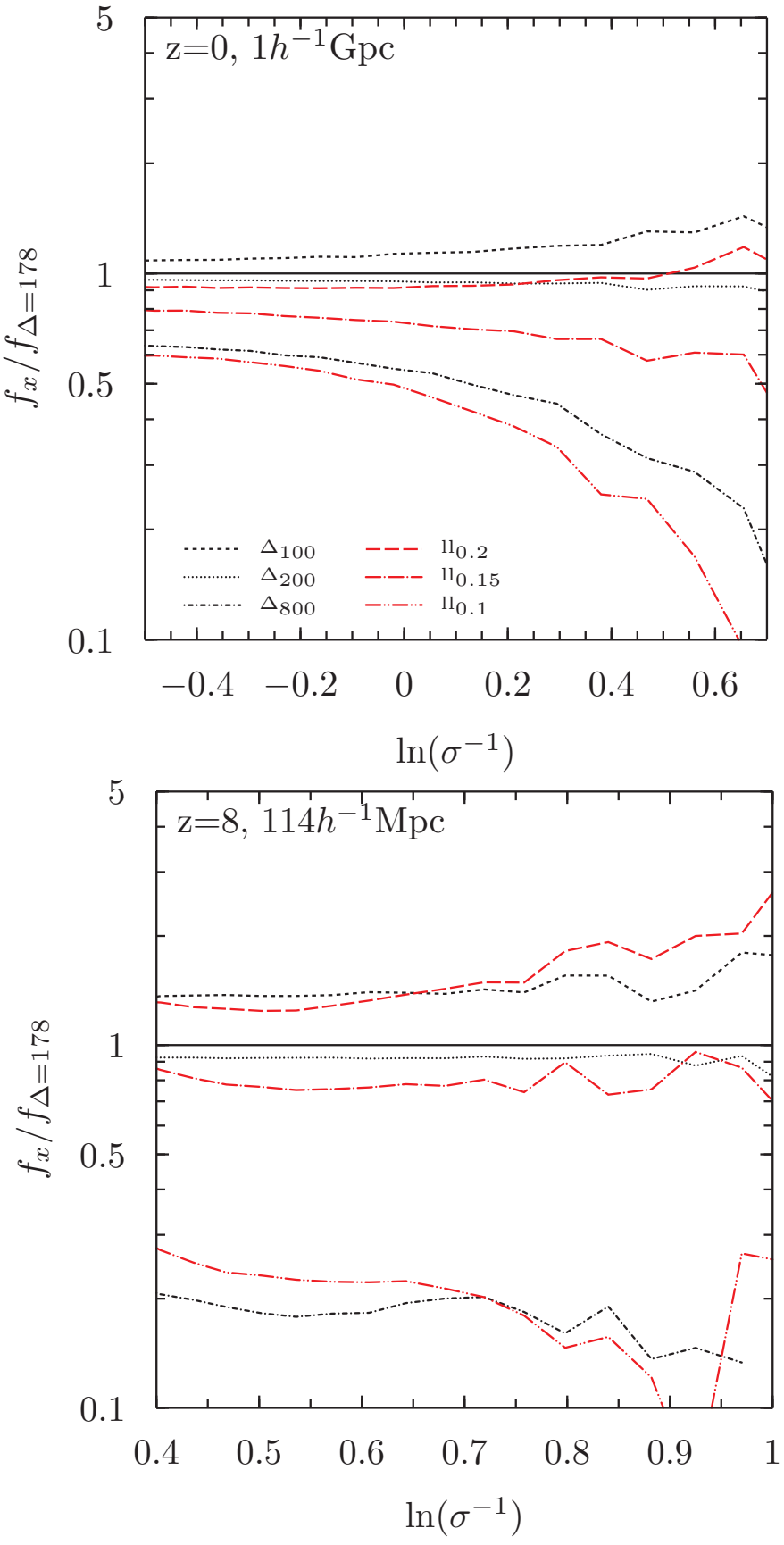


Figure 3.16: Comparison of mass functions from the 1 h^{-1}Gpc simulation at $z = 0$ (top panel) and the 114 h^{-1}Mpc simulation at $z = 8$ (bottom panel) using AHF with a variable overdensity and FOF with a variable linking length. All ratios are plotted on a base of the $\Delta = 178$ results.

redshift dependence (Courtin et al. (2011) via $\Delta(z)$, and the others via the concentration parameter, $c(z)$). Whilst the More et al. (2011) study looked at haloes up to $z = 2.5$, further study needs to be undertaken to investigate the suitability of existing relations between linking length and overdensity at higher redshifts.

3.4.4.3 SO vs. FOF mass functions

The differences we observed above between haloes found using the two approaches filter through to secondary results derived from halo catalogues, including the mass function. Figure 3.8 shows the systematic difference between the AHF ($\Delta = 178$) and the FOF ($b = 0.2$) mass functions over a wide range of redshifts. There is a close similarity between the two mass functions at $z = 0$, especially for lower mass haloes, with the collapsed fraction of mass in higher-mass haloes greater in the FOF case. Gradually, as we move to higher redshifts, we see the amplitude of the SO mass function fall to around 80% of the FOF mass function at lower masses, with a much more marked decrease at higher masses. Past redshifts of around $z \sim 6$ the high mass tail of the SO mass function is $\sim 50\%$ lower than the FOF mass function. The causes of the difference between the two mass functions can be summarised as contributions from (1) the relationship between the masses of a given SO halo and its FOF counterpart, which depend on the choices of b and Δ and also on the concentration parameter of the haloes; (2) the amount of overlinking of FOF haloes; (3) the relative mass difference in the two halo types that arises from pseudo mass evolution, as discussed in § 3.2.1 and (4) other systematic effects, including the SO algorithm not correctly interpreting the properties of non-spherical haloes and both algorithms failing to reliably describe merging systems. White (2001), White (2002), Lukic et al. (2009) and More et al. (2011) have addressed the first factor in detail. Prada et al. (2012) provide a detailed exposition of the evolution of the concentration parameter over redshift. Their findings illustrate that halo concentrations lie on a characteristic 'U' shape in the c-lnσ^{-1} plane. This shape exhibits modest evolution in redshift, with the concentration of the minimum becoming slightly smaller at higher redshifts. Davis et al. (1985); Ma & Bertschinger (1995); Cole & Lacey (1996); Lukic et al. (2009) have investigated the second and, to some extent, the fourth factors. The third factor has not been studied in detail and remains a topic for future study.

Chapter 4

Extreme objects in the Jubilee simulation

This chapter presents results from the Jubilee simulation, a large N-body, dark matter-only cosmological simulation with a volume of $V = (6\ h^{-1}\mathrm{Gpc})^3$, containing 6000^3 particles. The simulation volume is sufficient to probe extremely large length scales in the universe, whilst at the same time the particle count is high enough so that dark matter haloes down to $1.5 \times 10^{12}\ h^{-1}\mathrm{M}_\odot$ can be resolved. The contents of this chapter feature in Watson et al. (2013a), which has been submitted to MNRAS for publication.

4.1 Introduction

Surveys mapping a substantial portion of the observable Universe (e.g. BOSS (Dawson et al., 2013), WiggleZ (Drinkwater et al., 2010), BigBoss (Schlegel et al., 2009), Pan-STARRS (Magnier et al., 2013), DES (Mohr et al., 2012), PAU (Benítez et al., 2009), LSST (LSST Dark Energy Science Collaboration, 2012), Euclid (Amiaux et al., 2012), etc.) aim to constrain the cosmological model to unprecedented accuracy. As they will be able to capture very faint objects, they will have a shot-noise level low enough to be close to sampling variance-limited. This requires impressive handling of every step of the observational pipeline in order to limit the possibility of systematic errors that may degrade the information contained in them. Aside from these observational efforts, there will also be a similarly high demand placed on our ability to generate theoretical predictions that are equally accurate. This undoubtedly calls for numerical simulations of cosmic structure formation that resolve galactic scales in volumes comparable to the ones covered by these surveys. This is a weighty undertaking. A simulation must cover a wide

dynamic-range in order to accurately sample large-scale structure (LSS) in the universe. In particular, simulations need to resolve dark matter haloes, which are believed to host the observed galaxies, groups and clusters of galaxies, and accurately model the physics of galaxy formation and other non-linear physics, whilst adequately sampling large-scale matter fluctuations. Only recently have such simulations become feasible and nowadays full-box simulations of considerable fractions of the observable Universe are being conducted utilising close to a trillion particles (for a review of dark matter N-body simulations see Kuhlen et al. (2012)).

Whilst a careful comparison of the statistical clustering properties of objects, in particular galaxies, will put tighter constraints on the parameters of any cosmological model, it is worth noting that the mere existence of individual outliers might pose challenges: following observations of a series of apparently extreme objects (i.e. huge galaxy clusters at high-redshift such as XMMU J2235.3-2557 – a cluster with mass $\mathrm{M} > 4 \times 10^{14}\ h^{-1}\mathrm{M}_{\odot}$ at redshift $z \approx 1.4$ (Mullis et al., 2005; Rosati et al., 2009), some authors have claimed that such objects are highly unlikely to exist in a concordance ΛCDM cosmology and hence pose a challenge to its validity (Jimenez & Verde, 2009; Crocce et al., 2010; Lee & Komatsu, 2010; Cayón et al., 2011; Harrison & Coles, 2011; Hoyle et al., 2011; Waizmann et al., 2012; Holz & Perlmutter, 2012). This naturally leads to attempts to better understand the statistics of such rare objects but this again requires simulations of large enough volumes and sufficient resolution to properly capture the likelihood of their formation. Theoretical (as opposed to numerical) studies of such objects are challenging too since, due to their rarity and highly non-linear nature, they typically do not obey simple Gaussian statistics (Harrison & Coles, 2011, 2012).

In this chapter we present one of the largest cosmological dark matter only simulations to date, the so-called Juropa Hubble Volume, 'Jubilee' simulation[1], consisting of 6000^3 particles in a cubical volume of side-length 6 h^{-1}Gpc. It is laid out as follows. In § 4.2 we outline our methodology for running the simulation and deriving from it results including halo and void catalogues. In § 4.3 we present our main results and in § 4.4 briefly discuss their potential implications.

[1] http://jubilee-project.org

4.2 Methodology

4.2.1 The Jubilee simulation

The results presented in this chapter are based on a large-scale structure N-body simulation: the Jubilee simulation, which has 6000^3 (216 billion) particles in a volume of $6\,h^{-1}$Gpc. The particle mass is $7.49 \times 10^{10}\ h^{-1}\mathrm{M}_\odot$, yielding a minimum resolved halo mass (with 20 particles) of $1.49 \times 10^{12}\ h^{-1}\mathrm{M}_\odot$, corresponding to galaxies slightly more massive than the Milky Way. Luminous Red Galaxies (LRGs; $\mathrm{M} \sim 10^{13}\ h^{-1}\mathrm{M}_\odot$) are resolved with 100 particles, and galaxy clusters ($\mathrm{M} > 10^{14}\ h^{-1}\mathrm{M}_\odot$) are resolved with 10^3 particles or more. The simulation and most analyses were performed on the Juropa supercomputer at Jülich Supercomputing Centre in Germany (17,664 cores, 53 TB RAM, 207 TFlops peak performance) and required approximately 1.5 million core-hours to run. The simulation was run on 8,000 computing cores (1,000 MPI processes, each with 8 OpenMP threads).

4.2.2 Halo finding

We employ three halo finding codes in our analysis, as described in § 2.5: CPMSO, AHF and FOF. At $z = 0$ we identify over 400 million haloes and the first haloes in the simulation form at $z = 11$. The halofinding methods employed are as outlined in § 2.5.

4.2.3 Void finding

The formation of structure in the universe is a hierarchical process: small objects form, grow by accretion and merging and form more and more massive objects up to clusters of galaxies. Between the clusters large filaments can be seen both in observational data as well as in numerical simulations (Figures 2.2, 2.3 and 4.1). These filaments surround large regions of low density which do not contain objects as massive as the ones found in the filaments or the knots at the end of filaments. These low density regions – voids – are the most extended objects in the universe. There are many different ways to define voids and correspondingly there are many different void-finding algorithms (for a review, see Colberg et al., 2008). In the following we are interested in the largest spherical regions of the universe which do not contain any object above a certain threshold in mass. In principle, one could extend this definition of voids to non-spherical regions, however, in this case one can get arbitrary volumes depending on the shape allowed. Since we are interested in the void function, we restrict ourselves to spherical voids which are described

only by one parameter, their radius. We identify voids in a sample of point-like objects distributed in space. Here, these objects are our AHF haloes above a certain mass but one could also use galaxies above a certain luminosity. Thus, our voids are characterised by a threshold mass. If one decreases this mass threshold, the number of objects increases and the size of the void decreases. In fact, a given void defined with objects at a higher mass becomes decomposed into many smaller voids defined in the distribution of lower mass objects (Gottlöber et al., 2003). This reflects the scale-free nature of structure formation.

The void selection algorithm (detailed in Gottlöber et al., 2003) is similar to Einasto et al. (1989) and begins by constructing the minimal spanning tree for haloes with masses equal to, or above, the selected threshold. The algorithm then searches for the point in the simulation box which has the largest distance to this set of haloes. This is the centre of the largest void, which, after it is identified, is then excluded, so the search can be repeated to find the second largest void and so on. In principle, this algorithm allows for the construction of voids with arbitrary shape.

4.3 Results

4.3.1 Large-scale structure

In Figure 4.1 we show a slice of the Cosmic Web at $z = 0$ extracted from the Jubilee simulation. Perhaps most striking is the homogeneity of the matter distribution at large scales. This is expected from the cosmological principle discussed in § 2.1.1. On smaller scales significant non-linearities in the density field can be observed including voids, walls, filaments and clusters. For our observation of the ISW signal in Chapter 5 (and also in the creation of weak lensing maps and sky-catalogues of LRGs and radio galaxies), we place an observer inside the simulation volume at the very centre of the box. As can be seen in Figure 4.1, this is an unexceptional point in the density field and the full sky as observed by this central observer will show a highly homogeneous distribution of galaxies past a proper distance of a few hundred Mpc (i.e. a redshift of around $z \sim 0.1$).

In Fig. 4.2 we show the evolution of the power spectra of the density field, $P(k)$, from redshift $z = 6$ to $z = 0$. The particles were interpolated onto a regular grid of $12,000^3$ cells using the cloud-in-cell (CIC) interpolation scheme. From these data we then applied a correction for aliasing and the CIC window function and another for the effect of Poisson noise, all based on the prescription laid out in Jing (2005).

The baryonic acoustic oscillations (BAO) scale, $k \sim 0.1\ h\mathrm{Mpc}^{-1}$, is well within the

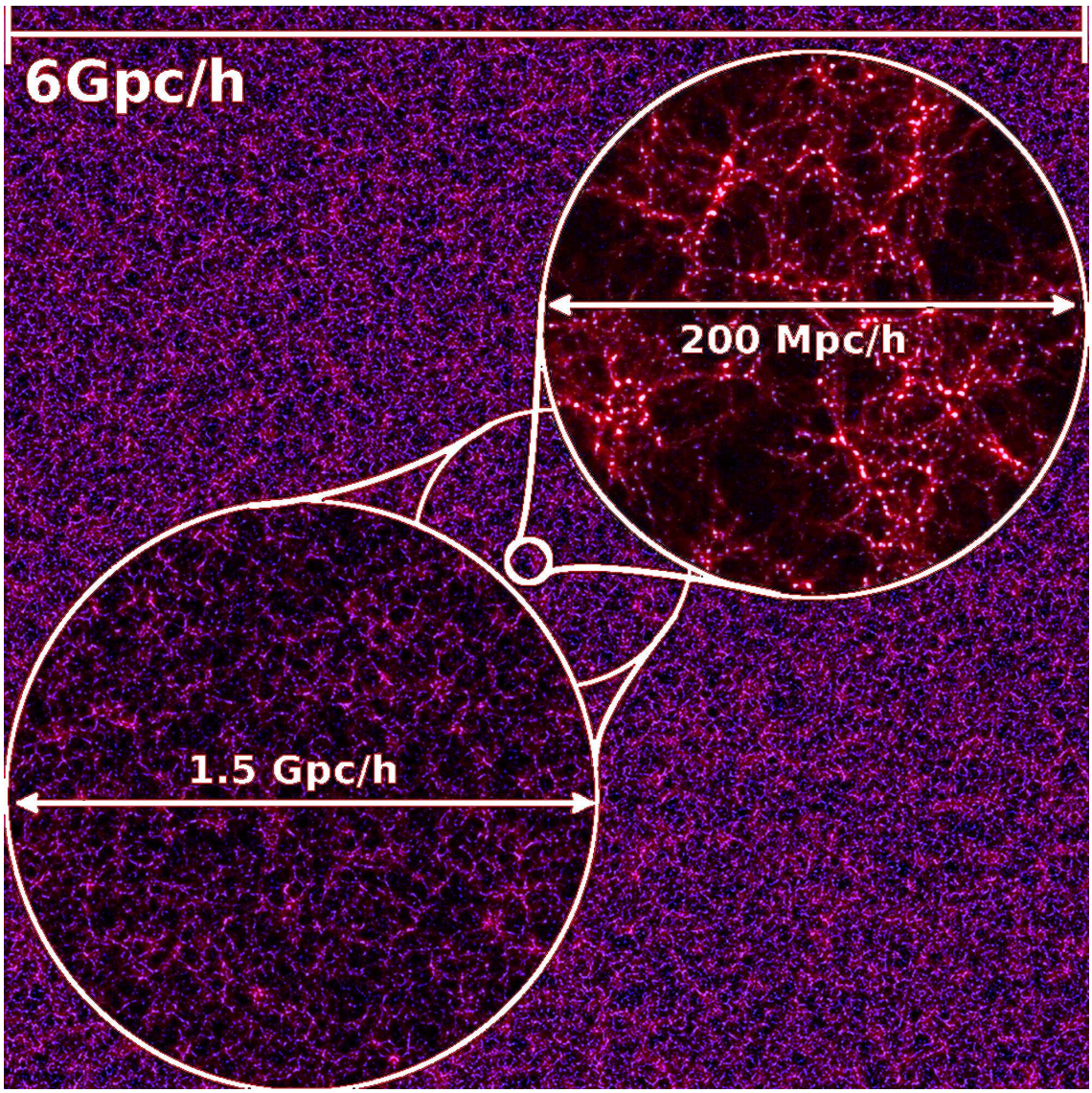


Figure 4.1: A slice of the Cosmic Web of structure at $z = 0$ based on the Jubilee simulation. The image is 6 h^{-1}Gpc per side and 20 Mpc thick.

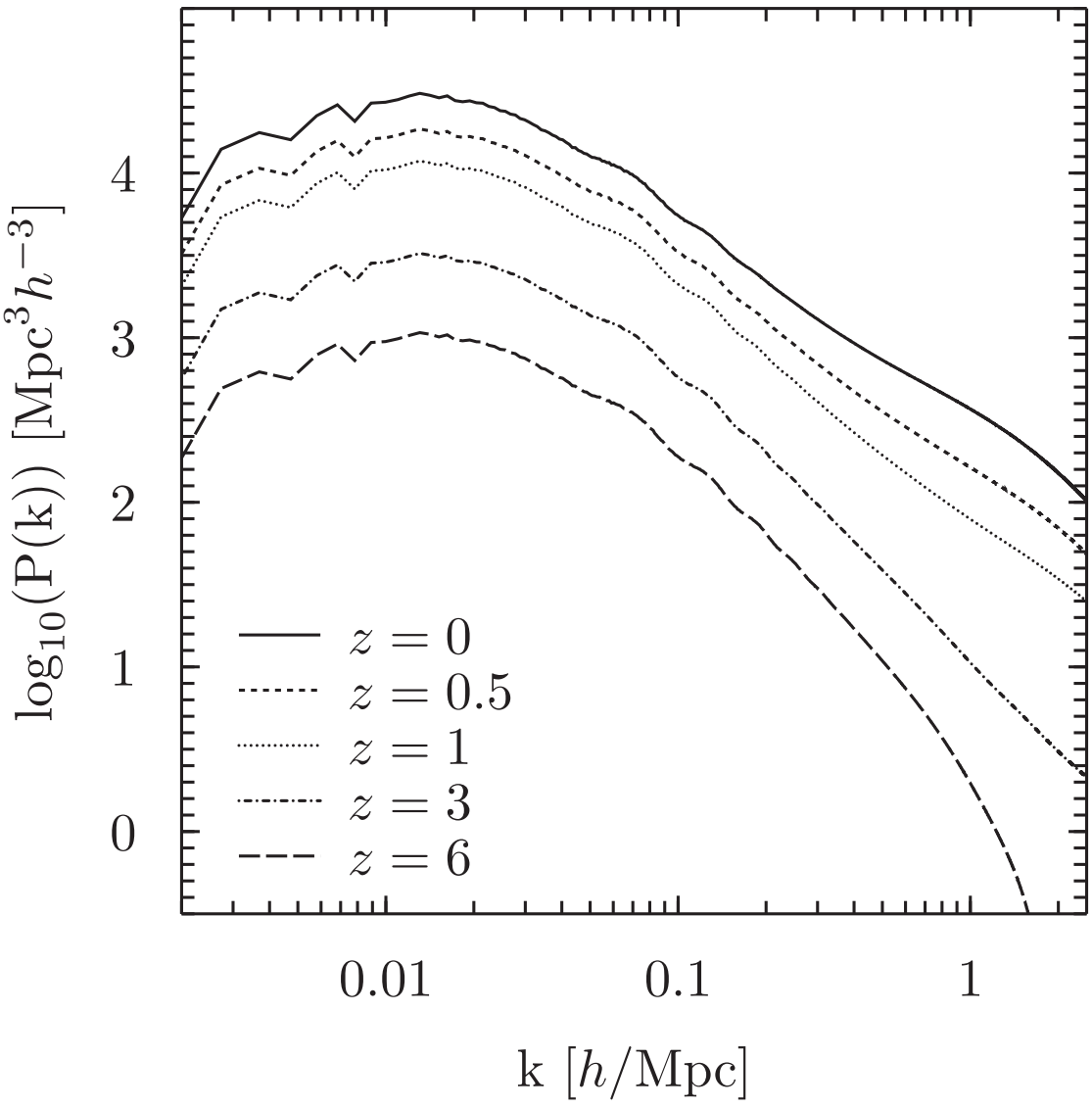


Figure 4.2: Evolution of the power spectrum of the density field, $P(k, z)$, as a function of the wavenumber, k, from the Jubilee simulation.

simulation box size and the BAO are clearly seen in the power spectra. At high redshift, $z \sim 6$, the power spectrum is largely linear, except at the smallest scales ($k > 1$ $h\mathrm{Mpc}^{-1}$). As the hierarchical structure formation proceeds, this non-linearity scale propagates to ever larger scales, reaching $k \sim 0.1$ $h\mathrm{Mpc}^{-1}$ at redshift $z = 0$, and thereby affecting the BAO scale.

We calculate halo mass functions using the three halofinding algorithms outlined in § 4.2.2. Figure 4.3 shows the residuals between our haloes and two fits from the literature at $z = 0$. We compare our CPMSO and AHF haloes to the Tinker et al. (2008) mass function, noting that the Tinker et al. (2008) fit was calibrated to haloes with an overdensity criteria of $\Delta = 200$ versus the background matter density and our haloes were calculated using

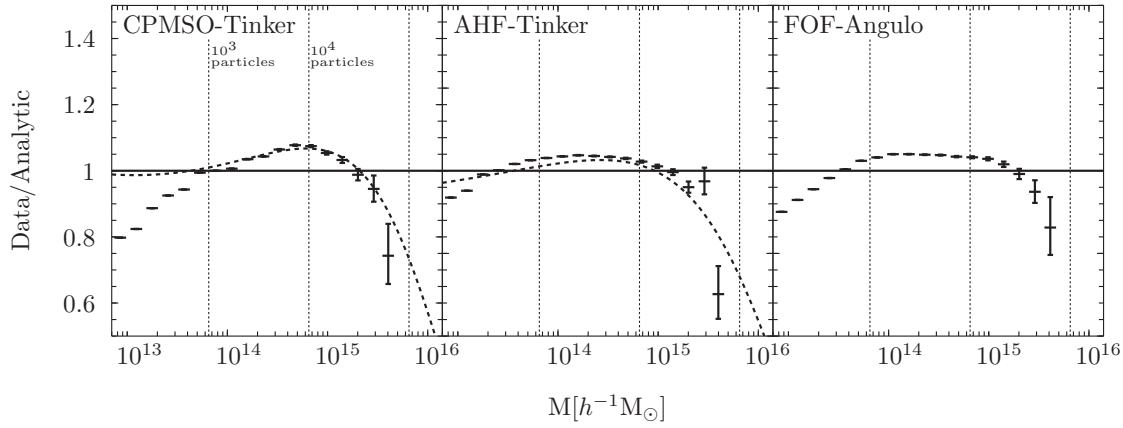


Figure 4.3: Jubilee simulation halo mass function based on different halo finders vs. recent analytic fits based on numerical data: (left panel) CPMSO vs. Tinker et al. (2008) fit, with the redshift-dependent fit from Chapter 3 (based on CPMSO haloes) shown as a dashed line; (middle panel) AHF vs. Tinker et al. (2008) fit, with the AHF $z = 0$ fit from Chapter 3 shown as a dashed line; and (right panel) FOF vs. Angulo et al. (2012). Errors shown are Poisson.

a value of $\Delta = 178$. Despite this we see a good correspondence to within $\sim 5\%$ between the Tinker et al. (2008) fit and our AHF data for haloes with particle counts greater than ~ 300. For the very largest haloes there is evidence that the Tinker et al. (2008) fit may be over predicting the mass function, although this is where shot noise begins to severely affect number counts of objects. The CPMSO data follows a similar trend to the AHF data. We overlay on these plots two of the fits from Chapter 3, mass function results calibrated to data that included the Jubilee haloes presented here. The fit used for the left-hand panel is the redshift-dependent fit based on the CPMSO halo-finder, the fit for the right-hand panel a fit based on AHF results for $z = 0$ (see Chapter 3 for further details).

We compare the FOF results to those of the Millennium, Millennium-2 and Millennium-XXL simulations (Angulo et al., 2012), the latter containing 6720^3 particles in a box with length 3 h^{-1}Gpc. The FOF halo data shows agreement to within $\pm 5\%$ with the Angulo et al. (2012) fit for haloes with 300 or more particles. The FOF haloes are being compared with a linking length of 0.2, which makes the similarity between the mass functions a good test of the validity of the Jubilee halo distribution since this was the same choice made in Angulo et al. (2012).

4.3.2 Cosmic variance

Due to the large size of our simulated volume we are able to quantify cosmic variance on scales smaller than our box size in terms of the number counts of objects one expects to find in a given volume. To that end we have compared halo counts in different mass bins in different sized subvolumes. We chose the subvolumes such that they filled the entire full-box with no overlap. The results are shown in Figure 4.4. We show the 1σ error in the number counts of haloes by mass bin relative to our entire $(6\ h^{-1}\mathrm{Gpc})^3$ volume for sub-box lengths of 3, 2, 1 and 0.5 h^{-1}Gpc. These choices directly compare to the box lengths of some contemporary simulations (Millennium-XXL (Angulo et al., 2012), Horizon (Teyssier et al., 2009), MultiDark (Prada et al., 2012) and Millennium (Springel et al., 2005) respectively). We also show a prediction for this error calculated by assuming that the halo number counts follow the AHF redshift parameterised mass function from Chapter 3, and by assuming that the observed error in number counts follows a Poisson distribution. This theoretical prediction matches the high-mass data very well. For lower masses the error becomes dominated by sample variance, as discussed in Smith & Marian (2011), and the Poisson prediction presented here begins to break down. We discuss how well the Poisson distribution matches the counts of high mass clusters in § 4.3.3 below.

As expected the error is minimal for lower mass haloes and increases for rarer objects. At $z = 0$ the 0.5 h^{-1}Gpc box has an error of under 10% up until haloes of mass around $4\times10^{14}\ h^{-1}\mathrm{M}_\odot$, while for box lengths of 1, 2 and 3 h^{-1}Gpc, a 10% error in number counts per mass interval is realised at around $1\times10^{15}, 2\times10^{15}$ and $3\times10^{15}\ h^{-1}\mathrm{M}_\odot$ respectively. The errors are exacerbated at higher redshifts, due to the haloes of a fixed size growing more rare at earlier times.

One subtlety should be mentioned: because the sub-volumes considered in this analysis were derived from a larger simulation volume they include the effect of matter fluctuations that exist on scales larger than their box lengths. We stress here that this is not the case for simulations with equivalent volumes to these sub-volumes, as modes of power in the density field that are larger than the box length of a simulation are typically set to zero. This implies that the variation in number counts presented here is slightly different from one that occurs due to a lack of appropriate large-scale power in a simulation volume. This mis-representation of reality (by all simulations, including the Jubilee despite its large volume) leads to an additional set of errors but is, fortuitously, only an issue for very small volume simulations with box lengths of the order of up to a few tens of Mpc (for example see Yoshida et al. (2003); Barkana & Loeb (2004); Sirko (2005); Power & Knebe

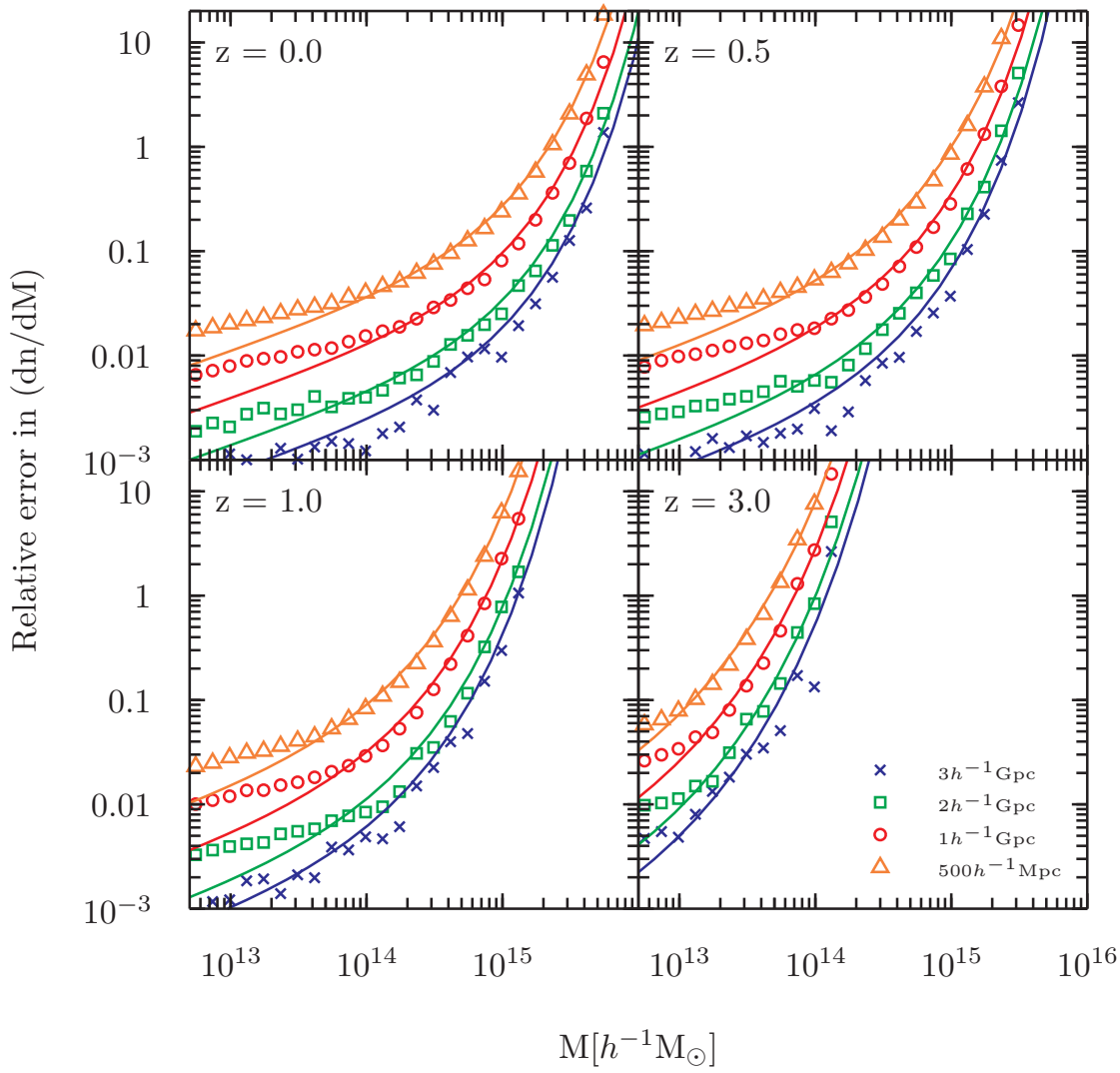


Figure 4.4: Relative error in dn/dM for subvolumes of box length 3, 2, 1, and 0.5 h^{-1}Gpc. 'Relative error' is defined as the standard deviation from the expected number counts per unit volume in a given mass interval, as defined using the entire 6 h^{-1}Gpc volume. The solid lines are the predictions based on the AHF redshift parameterised mass function from Chapter 3 combined with the assumption that the error in number counts around the mean is given by the Poisson distribution.

(2006); Bagla & Prasad (2006); Lukic et al. (2007)). Observational volumes, sampling the universe, do not suffer from this effect and the results presented here can be expected to translate reasonably well into counts of high-mass objects in LSS surveys.

4.3.3 Statistics of rare objects

A current topic in cosmology that relates to the number counts of very high mass objects is that of whether large observed clusters are in conflict with the standard ΛCDM model. In Figure 4.5 we show a theoretical prediction for the expected distribution of maximal mass clusters. This prediction was created using the Extreme Value Statistics prescription of Harrison & Coles (2011) and the redshift parameterised CPMSO mass function, adjusted for AHF haloes, from Chapter 3. It is comparable to the plot in Figure 1 of Harrison & Coles (2012) which was created using the mass function from Tinker et al. (2008), except for the fact that the mass of the haloes in this version is taken to be set by the $\Delta = 178$ overdensity criterion, rather than the $\Delta = 200$ criterion used in Harrison & Coles (2012). The black data points shown on Figure 4.5 correspond to the largest clusters observed in the Jubilee simulation by a central observer, also based on a $\Delta = 178$, as per the configuration of the CPMSO halofinder. The data all lies within the 3σ range showing the expected result that the there is no tension between objects observed in a ΛCDM cosmological simulation and the theoretical expectation from Extreme Value Statistics. Of interest is whether there are observed clusters in the Universe that have masses that are in tension with the ΛCDM model. To date observations have shown this to not be the case, as shown in a systematic review by Harrison & Hotchkiss (2012).

The question of how well the Poisson distribution fits our rare cluster number counts is addressed in Figure 4.6. The simulation volume at $z = 0.05$ was split up into 5438 independent subvolumes. For each subvolume we calculated the number of objects above a given threshold mass ($1.2 \times 10^{15}\ h^{-1}\mathrm{M}_\odot$, $1.4 \times 10^{15}\ h^{-1}\mathrm{M}_\odot$ and $1.6 \times 10^{15}\ h^{-1}\mathrm{M}_\odot$ for the panels in Figure 4.6, left-to-right respectively). The mass thresholds were chosen so that only a very small number (around 0–2) of objects were found in each subvolume, which represents the regime where we expect Poisson statistics to be dominant. We then compared the histogram of the measured distribution of the objects in the simulation to that predicted by a Poisson distribution with a mean set by the average across all the subvolumes. The correspondence between the two is very close. This is an interesting result as it validates the common choice of Poisson statistics for describing the expected distribution of these objects and is the first time it has been validated using a simulation

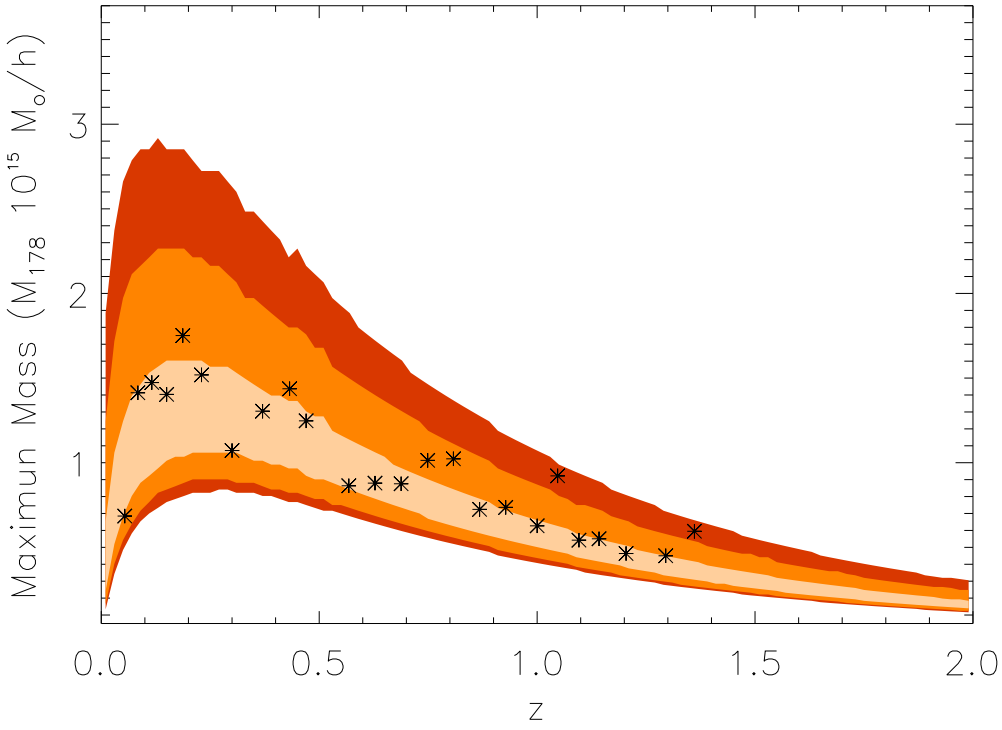


Figure 4.5: A new version of Figure 1 from Harrison & Coles (2012) using the Extreme-Value-Statistics prescription of Harrison & Coles (2011) and the CPMSO mass function from Chapter 3. The shaded regions show the 66%, 95% and 99% confidence intervals. The black data points show maximal mass clusters observed by a central observer in the Jubilee simulation.

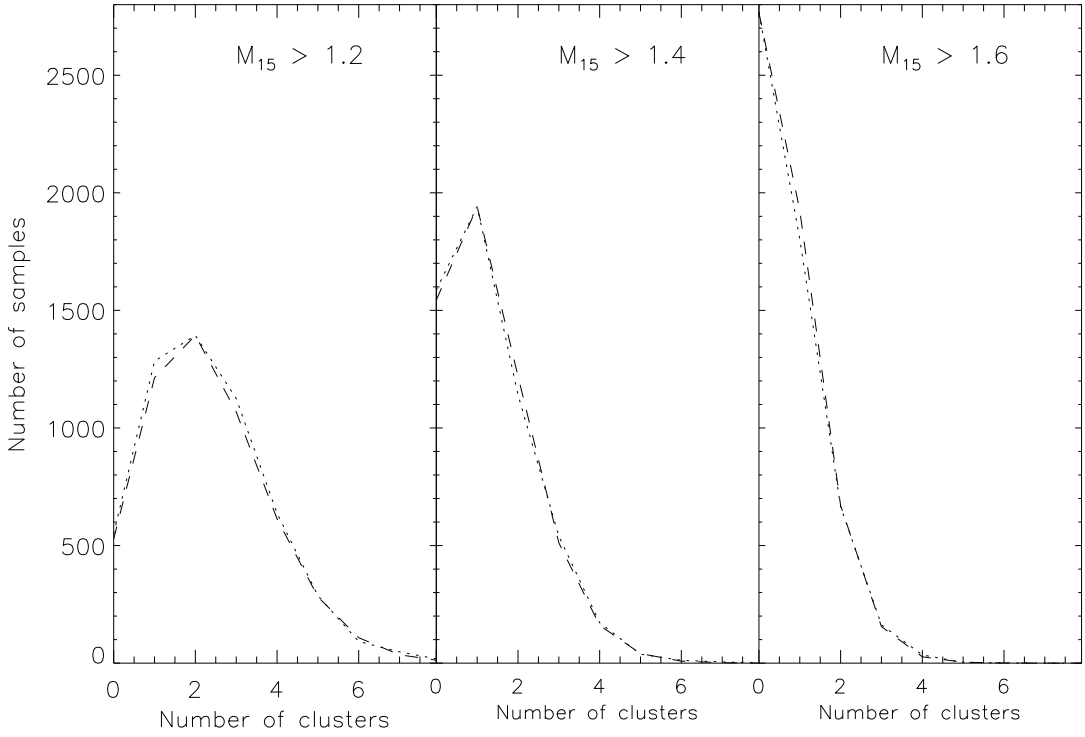


Figure 4.6: Histograms of extreme objects for three different thresholds in mass (dotted line) compared with the prediction from the Poisson distribution (dashed line) for the corresponding mean value of the objects above the corresponding threshold. The mass thresholds are: $1.2 \times 10^{15}\ h^{-1}\mathrm{M}_{\odot}$, $1.4 \times 10^{15}\ h^{-1}\mathrm{M}_{\odot}$ and $1.6 \times 10^{15}\ h^{-1}\mathrm{M}_{\odot}$ left-to-right panels respectively. The statistics are calculated for $z = 0.05$.

of this scale (for a detailed investigation of the applicability of the Poisson distribution in cluster counts across different masses see Smith & Marian (2011) who used simulations of box length 1.5 h^{-1}Gpc for their study).

4.3.4 High Δv mergers and the Bullet Cluster

There has been recent debate regarding whether the Bullet Cluster (1E0657-56, which resides at a redshift of $z = 0.296$) poses a challenge to the ΛCDM model. 1E0657-56 consists of a large cluster of mass $\mathrm{M}_{200} \sim 1.5 \times 10^{15}\ h^{-1}\mathrm{M}_{\odot}$ and a sub-cluster – the 'bullet' – of mass $\mathrm{M}_{200} \sim 1.5 \times 10^{14}\ h^{-1}\mathrm{M}_{\odot}$ that has traversed through the larger cluster, creating a substantial bow shock along the way (Markevitch et al., 2002; Barrena et al., 2002; Clowe et al., 2004, 2006; Bradač et al., 2006). Tension with ΛCDM arises from the calculated value for the speed of the shock of $v_s = 4740^{+710}_{-550}$ kms^{-1} (Markevitch et al., 2002;

Springel & Farrar, 2007), which was originally calculated by various authors (Hayashi & White, 2006; Farrar & Rosen, 2007) to be at the high end – perhaps too high – for a ΛCDM universe – whereas it might be better accommodated in alternative cosmologies (e.g. Llinares et al. (2009)). An important clarification of this issue was presented by groups working on simulations of Bullet-like systems (Takizawa, 2005, 2006; Milosavljević et al., 2007; Springel & Farrar, 2007; Mastropietro & Burkert, 2008) where, in general, it was found that the shock speed was substantially higher than the speed of the mass centroid of the infalling subcluster. For example, Springel & Farrar (2007) found that a Bullet-like system in their simulations had a shock speed of ~ 4500 kms^{-1} whereas the sub-cluster had a speed of only ~ 2600 kms^{-1}. Milosavljević et al. (2007) found that in an illustrative simulation the sub-cluster CDM halo had a speed that was 16% lower than that of the shock.

Even given this moderation of the extreme sub-cluster speed in 1E0657-56, there have still been claims in the literature that the ΛCDM model may be incapable of creating such a system (Lee & Komatsu, 2010; Thompson & Nagamine, 2012). This is not wholly unexpected as a) Mastropietro & Burkert (2008) have shown that the properties of the bow shock are not well described by simulations and b) even with a moderation in sub-cluster speed along the lines of Springel & Farrar (2007) or Milosavljević et al. (2007), the speed may still be too high for the ΛCDM model to accommodate. These studies have relied on numerical simulations to observe the distribution of relative velocities in colliding clusters. From these distributions 1E0657-56 can be assessed and deemed to be either rare for a ΛCDM universe or so rare that it puts the whole model in doubt.

Alternative approaches have also been taken in addressing this question. Forero-Romero et al. (2010) looked in 2-D-projected position-space for Bullet-like systems in the MareNostrum Universe, a large hydrodynamical cosmological simulation. The characteristic distribution of gas and dark matter in 1E0657-56, as projected on the sky – with a large displacement between the cluster's gas and dark matter – was found to be expected in 1% - 2% of clusters with masses larger than $10^{14}\ h^{-1}\mathrm{M}_{\odot}$. Nusser (2008) performed a 'back in time' analysis to place bounds on the relative overdensity in which the system resides, concluding that for a relative speed of ~ 4500 kms^{-1} the system would need to have a mass of $2.8 \times 10^{15}\ h^{-1}\mathrm{M}_{\odot}$ and exist in a local overdensity of 10 times the background density of the universe.

Here we bring to the debate the huge number counts of clusters in the Jubilee simulation. We consider clusters with mass greater than $7 \times 10^{13}\ h^{-1}\mathrm{M}_{\odot}$ that are colliding with

other haloes of equal or higher mass at $z = 0.3$. Our results are shown in Figure 4.7, along with the original Bullet speed presented in Markevitch et al. (2004), and the moderated result from Springel & Farrar (2007), which represents the lowest value from the literature to-date. As can be seen from the distribution the Bullet cluster is an extreme object. We find no mergers in our volume with a collision speed that equals or exceeds even the more conservative speed estimate for the cluster. This result is in line with previous attempts to use large cosmological simulations to address this issue (Hayashi & White, 2006; Lee & Komatsu, 2010; Thompson & Nagamine, 2012). Interestingly, Thompson & Nagamine (2012) extrapolated their results from smaller simulation volumes and concluded that a volume of $(4.5\ h^{-1}\mathrm{Gpc})^3$ would be required in order to observe a Bullet-like cluster. We have presented results here from a volume that is substantially larger and have been unable to observe such a cluster. We detect one merging system that comes close to the Bullet Cluster's relative pairwise velocities, with a speed of a little over 2000 kms^{-1} and a separation between haloes of $\sim 0.8\ h^{-1}\mathrm{Mpc}$. There are 34 other systems in the volume that have a merging velocity of over 1000 kms^{-1}, with a broad distribution of separations.

The conclusion that we tentatively put forward based on this result is that there is a tension between this result and the standard cosmological model. As we are not alone in reaching this conclusion, there would appear to be a need for careful further research into this question. However, we must also add a number of caveats to our result. First, we have not looked for the specific configuration of 1E0657-56 in terms of the masses of the host halo and its infalling sub-halo. Rather, we have merely taken a cut-off in mass that is appropriate for the lower mass halo upwards. Second, we have placed no restrictions on the directions of the relative velocities of the halo pairs. The bow shock observed in the Bullet Cluster has arisen from the Bullet sub-halo having passed through the parent halo (it is this occurrence, that, fortuitously, lies almost in the plane of the sky as we observe it, that has allowed us to identify the relative pairwise velocities of the halo and sub-halo in the system). In our analysis we plot *all* the pairwise velocities of the haloes, making no distinction between haloes that are infalling and haloes that have already undergone a collision nor considering how the orientations of the collisions might appear to a specific observer. This is a fair way to assess the data because the actual collision in a Bullet Cluster-like system is expected to take a few hundred Megayears and therefore is a relatively short event. Canvassing all our haloes in this manner assesses whether there is likely to be or whether there has been a Bullet Cluster-like collision in the simulation around $z \sim 0.3$. Had we detected a few clusters of comparable pairwise

velocity to the Bullet, we might then be in a position to ask how probable it would be for an observer to see the collision. As it is, given that we are searching in a volume of $(6\ h^{-1}\mathrm{Gpc})^3$ for a suitable collision and we find no Bullet-like candidates with equal to or higher pairwise velocities, it seems remarkable that the Bullet Cluster has been observed at all. Lastly, halo finding algorithms are notoriously sub-optimal when trying to find and separate haloes that are merging (this is discussed in detail in Knebe et al. (2011) – we draw the reader's attention to Figure 10 from that paper in particular). This has major implications for the detection of a Bullet Cluster in a simulation because the radial distance of the sub-halo from the parent cluster is observed to be only $0.6\ h^{-1}\mathrm{Mpc}$. Given that we use the CPMSO halofinder in this analysis, that does not identify sub-haloes within parents, it is possible, especially for small separations, that we are missing a population of merging haloes due to the mis-identification of haloes that are close to one another. Haloes that have collided, with mass centroids that are close to one another, are entirely omitted from our analysis and this is represented in Figure 4.7, where the population of merging clusters with small separations is very sparse.

4.3.5 The Jubilee void function

The distribution of voids for a given threshold is characterised by the void function, the number of spheres with radii larger than R_{void} per volume. We have studied the void distribution at redshifts $z = 0$, 0.5, and 1. At $z = 0$ we have identified the voids in the distribution of haloes more massive than $5 \times 10^{14}\ h^{-1}\mathrm{M}_{\odot}$, $2 \times 10^{14}\ h^{-1}\mathrm{M}_{\odot}$, $1 \times 10^{14}\ h^{-1}\mathrm{M}_{\odot}$, and $1 \times 10^{13}\ h^{-1}\mathrm{M}_{\odot}$. At redshift $z = 0$ we identified 244,989, 1,753,982, 5,596,627, 91,615,821 haloes more massive than these thresholds, respectively. Thus, the mean distance between them (i.e. the box length divided by the cube root of the number counts) is about 96 h^{-1}Mpc, 50 h^{-1}Mpc, 34 h^{-1}Mpc, and 13 h^{-1}Mpc. Nevertheless, we found huge volumes which do not contain any of these objects.

In Figure 4.8, top panel, we show the void functions at $z = 0$ for three different threshold masses. For the largest threshold, we find a few very large spheres with radii of 150 h^{-1}Mpc that do not contain any cluster more massive than $5 \times 10^{14}\ h^{-1}\mathrm{M}_{\odot}$. For smaller thresholds, the void function is very steep, i.e. there are a number of voids with a volume almost as large as the volume of the largest voids defined by the threshold. This means that the voids are almost uniformly distributed, as there are so many of a similar size. At higher redshifts (middle and bottom panel of Figure 4.8) we observe similar behaviour but, due to the evolution of the mass function, only with lower threshold

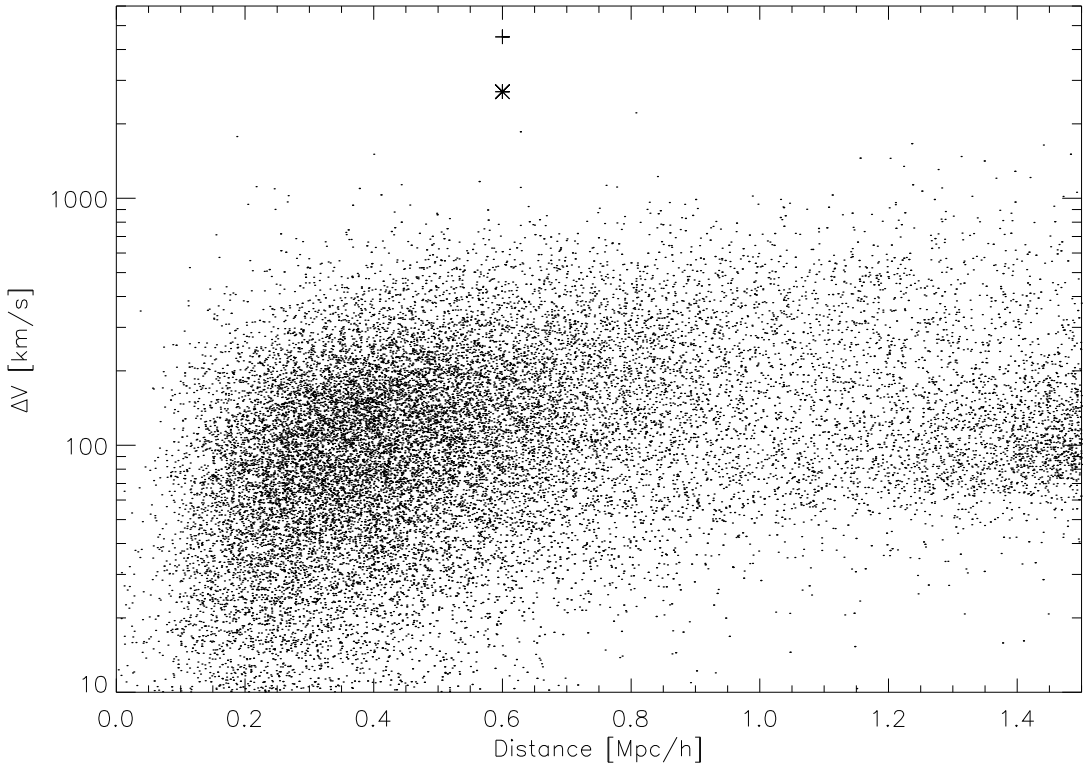


Figure 4.7: Relative pairwise velocities for haloes. The data points correspond to the observed bullet cluster speed from Markevitch et al. (2004) (plus symbol) and the corrected speed estimated by Springel & Farrar (2007) (asterisk symbol). The simulated speeds were obtained from the full box using the output redshift slice ($z = 0.3$) that best matched the redshift of the bullet cluster ($z = 0.296$). Only haloes with masses larger than $7 \times 10^{13}\ h^{-1}\mathrm{M}_{\odot}$ were considered in our search for colliding pairs.

masses. Note that, at the lowest threshold (10^{13} $h^{-1}\mathrm{M}_{\odot}$) the maximum void radius is almost redshift independent between $z = 0 - 1$ and occurs at a void radius of about 40 h^{-1}Mpc. This may seem in contradiction to the fact that low density regions expand slightly faster than the mean expansion rate of the universe. However, since the tracers of the voids are also evolving (i.e. the threshold mass increases with time), the number of objects above the threshold evolves. For 10^{13} $h^{-1}\mathrm{M}_{\odot}$ mass haloes, the number counts rise from 38,994,056 at $z = 1$ to 91,615,821 at $z = 0$. Therefore, the mean distance shrinks from 18 h^{-1}Mpc to 13 h^{-1}Mpc and using this threshold mass we see the interesting result that the maximum void radius remains almost constant in time.

These results clearly show the necessity for cosmologists to consider large volumes in the classic assumption that the universe is homogenous. Considering the very largest voids in the Jubilee box, at $z = 0$ we expect to see underdensities on scales up to 300 h^{-1}Mpc, depending on the threshold mass considered, although these voids are very rare. Our results show that in order to expect statistical homogeneity from a survey or simulation, a volume of $\sim (100 - 200\ h^{-1}\mathrm{Mpc})^3$ needs to be considered. This is in line with the observational results of previous authors (see, for example, Yadav et al. (2005); Sarkar et al. (2009); Sylos Labini & Baryshev (2010)) and complements our result on the effect of cosmic variance on number counts of massive objects.

4.4 Summary and discussion

In this chapter we have presented a broad range of results from a ΛCDM-based simulation. The results have focussed on rare objects, such as extremely massive clusters and large void regions. The simulation itself represents one of the largest undertaken to-date, with a volume of $(6\ h^{-1}\mathrm{Gpc})^3$ and haloes resolved down to 1.4×10^{12} $h^{-1}\mathrm{M}_{\odot}$, a resolution that allows the creation of mock LRG and cluster catalogues.

The distribution of dark matter haloes in the Jubilee was found to be well-described by fitting functions from the literature and these dark matter haloes have been used in chapter 3 to construct mass function fits across a broad range of redshifts and volumes. For the rare tail of the mass function, we have confirmed that the Poisson distribution describes well the number counts of objects. The masses of clusters with extremal masses in the Jubilee simulation were investigated across a range of redshifts and were found to agree well with both observation and theory, in particular the expected masses of the very largest objects found when using Extreme Value Statistics.

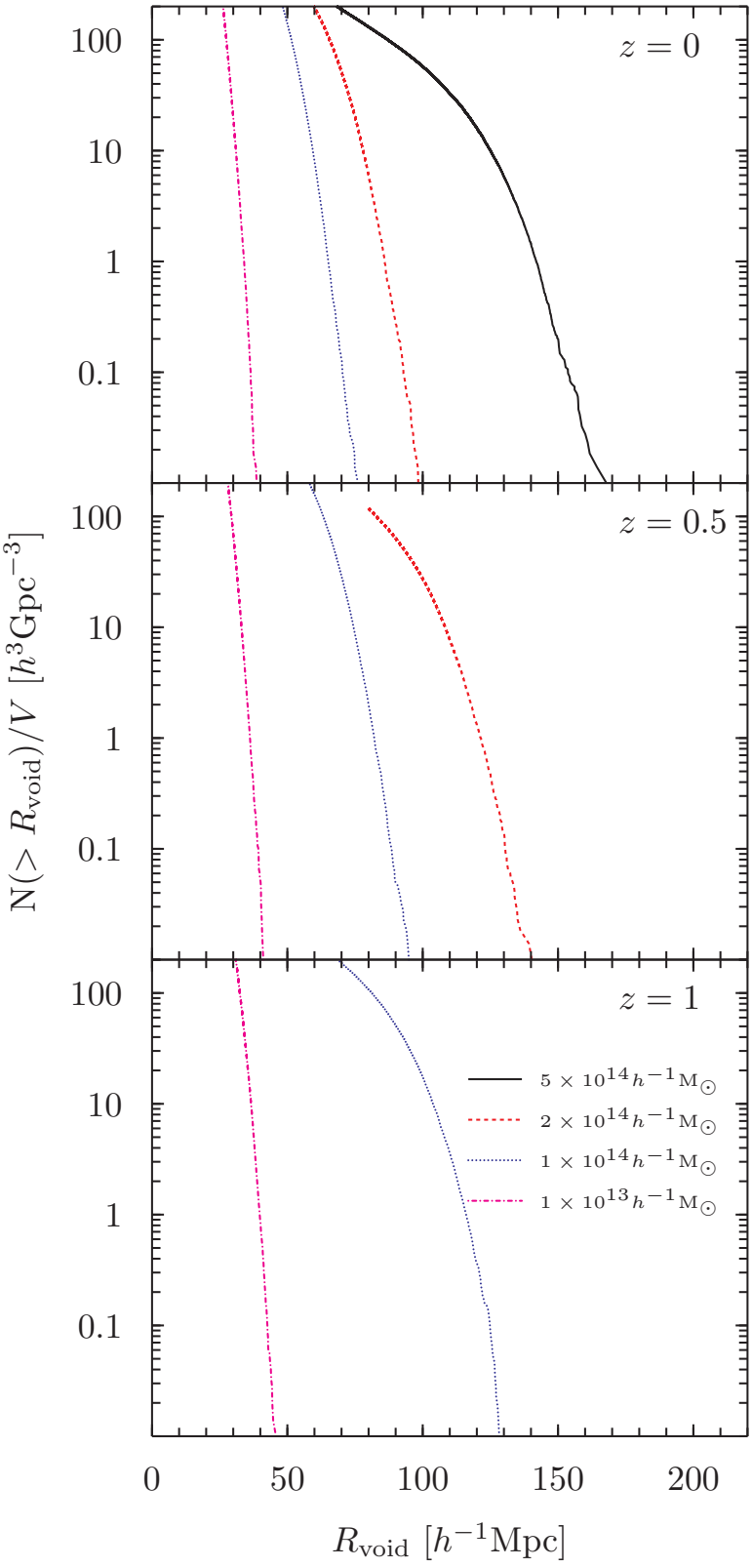


Figure 4.8: Void functions from the Jubilee simulation for $z = 0, 0.5, 1$ top-to-bottom respectively. Voids are defined as spherical regions of radius R_{void} wherein no haloes with a mass higher than a threshold mass are found. The plot shows, for different threshold masses, the number densities of voids with radii over R_{void}.

4.4.1 Implications for precision cosmology

An important prediction from this simulation is the expected effect of cosmic variance on the counts of massive clusters. This result can be used to gauge number-count errors in survey and simulation data. Understanding this is a vital component of the drive towards high precision cosmology. We showed in Figure 4.4 how the expected number of clusters in given volumes are likely to vary. In general, quantifying the effect of cosmic variance in simulations is notoriously tricky due to the requirement for either multiple repeats of a simulation or for a large simulated volume (or preferably both of these). Due to the large scale of the Jubilee volume we are able to use the latter and do so in a manner that includes the long-wavelength modes of the matter distribution. The variation in cluster counts for smaller boxes or surveys is highly significant if one is investigating the distribution of high-mass objects such as galaxy clusters, which form an important cosmological probe.

4.4.2 The largest voids

Our largest void, defined using a threshold mass of $5 \times 10^{14}\ h^{-1}\mathrm{M}_{\odot}$, is $\sim 350\ h^{-1}\mathrm{Mpc}$ across. This extreme void emphasises the need to carefully consider the scale at which the universe can be considered homogenous. To put this void in context, it is around one fifth of the volume of the Millennium simulation and it contains no clusters with mass greater than $5 \times 10^{14}\ h^{-1}\mathrm{M}_{\odot}$. The probability, based on volume-occupation alone, of finding yourself within this void in the universe represented by the Jubilee simulation is 0.01%. There have been investigations as to whether our occupying a local underdensity might explain the apparent existence of an accelerated expansion in the late-time universe (Ellis, 1979; Mustapha et al., 1997; Zehavi et al., 1998; Tomita, 2001; Iguchi et al., 2002; Barausse et al., 2005; Wiltshire, 2005; Moffat, 2005; Alexander et al., 2009; February et al., 2010; Marra & Pääkkönen, 2010; Nadathur & Sarkar, 2011). The void in question would need to have very specific characteristics that include its radius, sphericity, density and density profile. Predictions for these void parameters vary but have typically required the void to be of at least a few hundred Mpc in radius and, importantly, close to spherically symmetric, with us as observers very near its centre. This latter requirement is due to the type Ia supernovae data implying that dark energy is close to isotropic across the sky. We see from our void functions in Figure 4.8 that there are a few hundred voids in the Jubilee volume with radii $R_{void} > 100\ h^{-1}\mathrm{Mpc}$, for the $5 \times 10^{14}\ h^{-1}\mathrm{M}_{\odot}$ mass threshold. We estimate the proportion of the entire simulation volume taken up by voids with a radius of $R_{void} > 100\ h^{-1}\mathrm{Mpc}$ to be 0.04%. Adding an additional requirement that an observer

occupy the central 1% of the void volumes in question, we arrive at the total spatial volume in the Jubilee box that would contain observers in the centre of voids of radii greater than 100 h^{-1}Mpc to be $\sim 0.0004\%$. This is a rough estimate and ignores the fact that observers might be better considered to only exist at the locations of galaxies in the simulation. In addition, the simulation contains a dark energy component so has already modelled the effect of late-time accelerated expansion on structure formation. This latter point does not alter the order of magnitude of the result as void sizes in universes without dark energy are comparable to void sizes in ΛCDM (Müller et al., 2000).

4.4.3 The ΛCDM model versus observations

In general, the comparison of our results to observations implies that the ΛCDM model can be taken as a good model for describing the universe on cosmological scales. The distribution of most-massive clusters in the Jubilee was found to be in line with current observations and current theoretical predictions based on Extreme Value Statistics. The nature of Extreme Value Statistics is that it lacks predictive power in terms of constraining models but it is a powerful method for ruling out models based on only a handful of extreme data points. Had the masses of observed clusters in Figure 4.5 lain significantly away from the expected distribution, then the ΛCDM model would be immediately placed in doubt. One result in this chapter, that of the extreme nature of the Bullet cluster, is suggestive of a possible tension with ΛCDM. An Extreme Value Statistics approach is likely to cast this result in a more comprehensive context.

Chapter 5

The ISW effect and its cross-correlation with LSS in the Jubilee simulation

This chapter studies the ISW effect in the Jubilee simulation, as well as the construction of all-sky maps of weak lensing observables, mock catalogues of LRGs and the expected cross-correlation signal between the ISW and LSS. Much of the content of this chapter can be found in Watson et al. (2013b), which has been submitted for publication in MNRAS.

5.1 Introduction

The recent results from the Planck satellite (Planck Collaboration, 2013a) have shown the standard ΛCDM cosmological model to be in good health. The universe, as we currently understand it, consists mainly of some form of dark energy or cosmological constant (Λ) and a cold dark matter component. The key challenges in cosmology, however, remain the same: we still need to uncover the secrets of the dark sector. What is dark matter? What are the properties of dark energy?

To answer the latter question, the late-time integrated Sachs-Wolfe (ISW) effect (Sachs & Wolfe, 1967; Rees & Sciama, 1968; Hu & Sugiyama, 1994) can be a useful cosmological probe, since it is sensitive to the dynamical effects of dark energy and may thus be used to discriminate between different cosmological models (Crittenden & Turok, 1996; Afshordi et al., 2004). The effect is manifested as secondary anisotropies in the CMB radiation temperature, which are created when photons from the last scattering surface travel through time-evolving fluctuations in the gravitational potential, Φ, caused by LSS along their

path. For a flat universe filled entirely with a pressureless fluid, such as dark matter, Φ is constant with time at linear order, so that to first order the linear ISW effect is zero, although second order effects would arise, primarily due to the velocity field of the structures that seed the potential. The time evolution of Φ requires a significant non-pressureless component of the cosmological fluid (Sachs & Wolfe, 1967) or non-zero curvature (Kamionkowski & Spergel, 1994). Given that Planck shows the universe to be very close to flat (Planck Collaboration, 2013c), a detection of the ISW effect constitutes a direct measure of the effects of dark energy.

However, the detection of the ISW effect is complicated by two factors. The first is that the amplitude of the effect is an order of magnitude smaller than primordial anisotropies in the CMB. The second is that the ISW contribution to the CMB temperature power spectrum is greatest on large angular scales. This means that the detection is very susceptible to cosmic variance and also that measuring the signal through cross-correlation of CMB temperatures with LSS requires the use of galaxy surveys covering a large sky fraction and containing a very large number of galaxies (Afshordi et al., 2004; Douspis et al., 2008).

Following the earliest reported detections by Fosalba, Gaztañaga & Castander (2003); Boughn & Crittenden (2004); Afshordi, Loh & Strauss (2004); Nolta et al. (2004), most studies of the ISW effect have been based on a full cross-correlation between the CMB and different LSS catalogues that trace the matter density. Different techniques to achieve this calculate the cross-correlation in either real (e.g. Boughn & Crittenden, 2002; Giannantonio et al., 2008), harmonic (e.g. Afshordi et al., 2004; Schiavon et al., 2012) or wavelet (e.g. Vielva et al., 2006; McEwen et al., 2007) space. The results of these studies have been mixed, with reported detection significances ranging from $\sim 0-4\sigma$ (see Planck Collaboration, 2013b, for a recent study and a brief review of previous results). Recently, the Planck collaboration has also been able to cross-correlate the CMB map with a map of the reconstructed lensing potential, finding a $\sim 2.5\sigma$ significant detection of the ISW-lensing cross-correlation (Planck Collaboration, 2013b).

A different approach, using a stacking analysis of CMB patches along lines of sight that correspond to individual over- or underdensities identified in a galaxy survey, was found by Granett, Neyrinck & Szapudi (2008) to give a detection significance $> 4\sigma$, a result recently confirmed by Planck Collaboration (2013b) using the same lines of sight. The amplitude of the signal observed in this approach is, however, too large for the standard ΛCDM cosmology (Nadathur, Hotchkiss & Sarkar, 2012; Flender, Hotchkiss & Nadathur, 2013;

Hernandez-Monteagudo & Smith, 2012) and is currently unexplained. Subsequent stacking investigations using a different catalogue of voids have not shown the same strength of signal (Ilic, Langer & Douspis, 2013; Planck Collaboration, 2013b), adding to the mystery. Given the wide range of results and the uncertainties involved in their interpretations, a great deal of importance is placed on improving our theoretical understanding of the expected ISW effect in a ΛCDM cosmology. This may be best addressed by using large N-body simulations.

Whilst the large-scale ISW effect is governed by the dark energy-driven time variability of the gravitational potential – and is therefore observed in the radial direction – variations in the tangential direction of the potential results in achromatic path distortions of the photons (i.e with no gain or loss of energy). These tangential distortions are the gravitational lensing effect (see Hoekstra & Jain (2008) for a review). Lensing distortions concentrate on the small scales (of the order of a few arcminutes) and hence complement the large-scale ISW effect. The lensing effect does not depend (at least not to first order) on dark energy but is very sensitive to the distribution of the total mass. Due to this direct dependency on dark matter, gravitational lensing can produce reliable estimates of the matter power spectrum and thus provide independent and robust estimates of the cosmological model. Measurements of the CMB lensing effect can be used to set constraints on the spatial curvature, dark energy or neutrino masses (Mandelbaum et al., 2013) that are normally degenerate when only the CMB power spectrum is available.

Much simulation work has already been undertaken on the topic of weak lensing (see, for example, Barber et al., 1999; Jain et al., 2000; Vale & White, 2003; Carbone et al., 2008; Das & Bode, 2008; Fosalba et al., 2008; Hilbert et al., 2009; Teyssier et al., 2009; Lawrence et al., 2010; Kiessling et al., 2011; Carbone et al., 2013). However, simulations are typically based on boxes that are much smaller than the Hubble volume (with $L_{box} \sim 500\ h^{-1}$Mpc to 1 h^{-1}Gpc, although Teyssier et al. (2009) and Fosalba et al. (2008) consider boxes of length 2 and 3 h^{-1}Gpc respectively). For future surveys a much larger volume would be more suitable especially for the case of CMB lensing where the lensing cross section peaks at around $z = 1$ (i.e. around 2.3 h^{-1}Gpc).

In this chapter we study both the ISW and weak lensing effects using the Jubilee simulation. The simulation contains 6000^3 particles in a box of side 6 h^{-1}Gpc. It is therefore possible to use the simulation to model the ISW effect due to large-scale structure out to $z = 1.4$ without having to repeat the box (a shortcoming of previous, smaller, ISW simulations; see, for example, Cai et al., 2010, and the discussion in § 5.3.1.1, below).

Furthermore, with its high particle count we are able to directly resolve dark matter haloes that contain Luminous Red Galaxies (LRGs). This allows us to measure the cross-correlation between the simulated ISW and the large-scale structure traced by the LRGs on larger scales than has hitherto been possible. Direct measurement of the expected stacking signal from LRGs is also possible, as well as studies of the ISW-lensing cross-correlation.

This chapter details methodologies for the creation of mock LRGs, all-sky weak lensing maps and the ISW effect. It also presents results for the ISW-LSS cross-correlation signal. The chapter is laid out as follows. We first provide an overview of how the ISW maps, LRG catalogues and weak lensing maps were created in § 5.2. We then present the results from these modelling procedures and also the ISW-LSS cross-correlation signal in § 5.3. Finally, in § 5.4, we conclude with some general comments on the implications of this work for future ISW-detection efforts.

5.2 Methodology

5.2.1 The ISW effect in the Jubilee simulation

The ISW maps are produced adopting a semi-linear approach where the potential is computed exactly in the entire simulation box but its time derivative is computed using linear theory. In a recent work Cai et al. (2010) demonstrated that this approximation (hereinafter referred to as the LAV approximation, following the terminology of Cai et al.) is sufficient to study the ISW on the largest scales with indistinguishable results up to $\ell = 40$ in contrast to the exact (non-linear and computationally more expensive) calculation. At $\ell = 100$, the LAV approximation under-predicts the real power by nearly an order of magnitude since it does not account for the peculiar velocities that become important at small scales. Nevertheless, most of the ISW effect is concentrated on the largest scales ($\ell < 50$), for which the LAV is accurate. Because of this, and due to the maximum ISW-LSS cross-correlation signal being expected to occur around $\ell \sim 10$ for an LSS galaxy survey (Cooray, 2002), the use of the LAV in this work is appropriate.

The temperature fluctuations in the CMB induced by the ISW effect can be written as (Sachs & Wolfe, 1967):

$$\frac{\Delta T}{T} = \frac{2}{c^2} \int_{t_{CMB}}^{t_0} \dot{\Phi}(\mathbf{x}, t) dt, \tag{5.1}$$

where $\dot{\Phi}$ is the derivative of the gravitational potential with respect to time. The potential

can be calculated from fluctuations in the density field of the universe via the cosmological Poisson equation:

$$\nabla^2\Phi(\mathbf{x},t) = 4\pi G\rho_m(t)a^2(t)\delta(\mathbf{x},t), \tag{5.2}$$

where ρ_m is the background matter density and δ is the overdensity as defined by equation 2.40 (note that in this context the overdensity can be in the non-linear regime). In Fourier space equation 5.2 is

$$-k^2\Phi(\mathbf{k},t) = 4\pi G\rho_m(t)a^2(t)\delta(\mathbf{k},t). \tag{5.3}$$

Using the present day matter density parameter, $\Omega_{m0} = 8\pi G\rho_{m0}/3H_0^2$, and the fact that $\rho_m(t) = \rho_{m0}a(t)^{-3}$ (equation 2.12), we have

$$\Phi(\mathbf{k};t) = -\frac{3}{2}\Omega_{\rm m0}\frac{H_0^2}{k^2}\frac{\delta(\mathbf{k};t)}{a(t)}. \tag{5.4}$$

Differentiating this with respect to time then gives

$$\dot{\Phi}(\mathbf{k};t) = \frac{3}{2}\Omega_{\rm m0}\frac{H_0^2}{k^2}\left[\frac{H(t)}{a(t)}\delta(\mathbf{k};t) - \frac{\dot{\delta}(\mathbf{k};t)}{a(t)}\right]. \tag{5.5}$$

For the construction of the ISW maps we make the approximation that the velocity field is given by linear theory, where

$$\dot{\delta}(\mathbf{k};t) = \dot{D}(t)\ \delta(\mathbf{k};t=0), \tag{5.6}$$

and $D(t)$ is the growth factor used in equation 3.1 (Heath, 1977). We can substitute for $\dot{\delta}(\mathbf{k};t)$ in equation 5.5 resulting in

$$\dot{\Phi}(\mathbf{k};t) = \frac{3}{2}\Omega_{m0}\frac{H_0^2}{k^2}\frac{\delta(\mathbf{k},t)}{a(t)}H(t)\left(1-\beta(t)\right), \tag{5.7}$$

where $\beta(t) = d\ln D(t)/d\ln a(t)$. Finally, combining equations 5.4 and 5.7 results in

$$\dot{\Phi} = -\Phi H(t)[1-\beta(t)], \tag{5.8}$$

which is valid in both real and Fourier space.

To calculate the ISW effect in the Jubilee simulation, we first produce a smoothed overdensity field, $\delta(\mathbf{x}, t)$, from the particle outputs from 20 checkpoints between $z = 0$ to 1.4. The overdensity field is calculated using a Cloud-In-Cell (CIC) smoothing kernel (see, for example, Hockney & Eastwood, 1988). Then, from the $\delta(\mathbf{x}, t)$ field, we use the Multiple Fourier Transform (MFT) method (Hockney & Eastwood, 1988) to calculate the potential field $\Phi(\mathbf{k}, t)$. This follows the steps outlined above, solving the Poisson equation in the Fourier domain. We produce maps of the real-space potential in redshift shells given by the distribution of the simulation time slices, which totalled 20 between $z = 0$ to 1.4. To produce the maps, we traced rays from a centrally-located observer through each of the cells of the potential field. The potential for each shell, integrated along lines of sight in this manner, was then projected onto the sky using HEALPix[1] (Górski et al., 2005). We applied a linear interpolation between the different slices in order to account for potential values at intermediate redshifts (the net effect of interpolating versus not interpolating is $< 1\%$ on the final results). From these outputs we used equation 5.8 to calculate $\dot{\Phi}$ and calculated the ISW effect using equation 5.1.

The data handling requirements for analysing the Jubilee simulation in this manner were particularly challenging. For each output slice the simulation's particle data totalled around 4TB. These outputs were then analysed and converted into density and then potential fields. The mesh used for the potential fields was 6000^3 in size ($(1\ h^{-1}\mathrm{Mpc})^3$ per cell) so each output slice in redshift for the potentials was 800Gb in size. Overall, the data for the potential between $z = 0$ to 1.4 totalled over 15TB and was reduced from particle data that was 100TB in size. For the weak lensing outputs, discussed in § 5.2.3 below, five derivatives of the potential were calculated, resulting in another 75TB of data.

5.2.2 LRG catalogue construction

For correlating the ISW with LSS, we first need to create a suitable catalogue of tracers of the dark matter density field. For the ISW-LSS signal, as we shall see, a population of tracers that exist between redshifts of $z \sim 0.2$ to 0.8 create the strongest signal. LRGs are, therefore, very useful because they are detectable across the range in question due to their high luminosities. The majority of LRGs reside in haloes that have masses in excess of $10^{13}\ h^{-1}\mathrm{M}_\odot$ (Zheng et al., 2009). They are typically the Brightest Cluster Galaxy (BCG)

[1] http://healpix.jpl.nasa.gov

in their cluster and are located at the centre of their parent dark matter haloes (Zheng et al., 2009; Wen et al., 2012; Zitrin et al., 2012) (although note that the corollary is not true: BCGs are not typically LRGs: Wen et al. (2012) show that $\sim 25\%$ of BCGs are LRGs). Complications arise in high mass clusters where there exists a fraction of LRGs ($\sim 5\%$) that are satellites (Zheng et al., 2009). In this study we ignore satellite LRGs and model only a population of central LRGs in our dark matter haloes. To create an LRG catalogue we need to find dark matter haloes in our simulation. We used the CPMSO haloes (see § 2.5) for this.

5.2.2.1 Modelling of central LRGs in haloes

We applied part of a Halo Occupation Distribution (HOD) model to populate our haloes with LRGs. The model we used was that of Zheng et al. (2009) who studied a sample of LRGs from the Sloan Digital Sky Survey (Eisenstein et al., 2005) from $z = 0.16$ to 0.44. We apply, specifically, the prescription laid out in appendix B of Zheng et al. (2009) which gives the average occupation function (based on Zheng et al. (2005)) for *central* LRGs as

$$\langle \mathrm{N_{cen}} \rangle_{\mathrm{M}} = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\log\ \mathrm{M} - \log\ \mathrm{M_{min}}}{\sigma_{\log\ \mathrm{M}}}\right)\right], \tag{5.9}$$

where erf is the error function:

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt, \tag{5.10}$$

$\langle \mathrm{N_{cen}} \rangle_{\mathrm{M}}$ is the average number of central LRG galaxies in a halo of mass M, $\sigma_{\log\ \mathrm{M}}$ controls the width in the log M-N relation and $\mathrm{M_{min}}$ is a characteristic minimum mass of hosts with central galaxies. The central LRGs follow a nearest integer probability distribution, i.e. a halo is either occupied by an LRG or not, based on whether a random number between 0 and 1 lies closer to 0 (unoccupied halo) or 1 (occupied halo). This model allows us to populate central LRGs in our haloes using a random number generator.

The variables in equation 5.9 were calculated by Zheng et al. (2009), based on a volume-limited sample of LRGs with a redshift range of $z = 0.16$ to 0.44. The absolute magnitude cut off for this sample was based on a rest frame g-band magnitude of $M_g < -21.2$ (note that we refer to masses as unitalicised, M, and magnitudes as italicised, M) which was calculated at $z = 0.3$ for all LRGs and included corrections for evolution. The HOD parameters were found to be: $\log\ \mathrm{M_{min}} = 13.673 \pm 0.06\ h^{-1}\mathrm{M}_\odot$ and $\sigma_{\log \mathrm{M}} = 0.621 \pm$

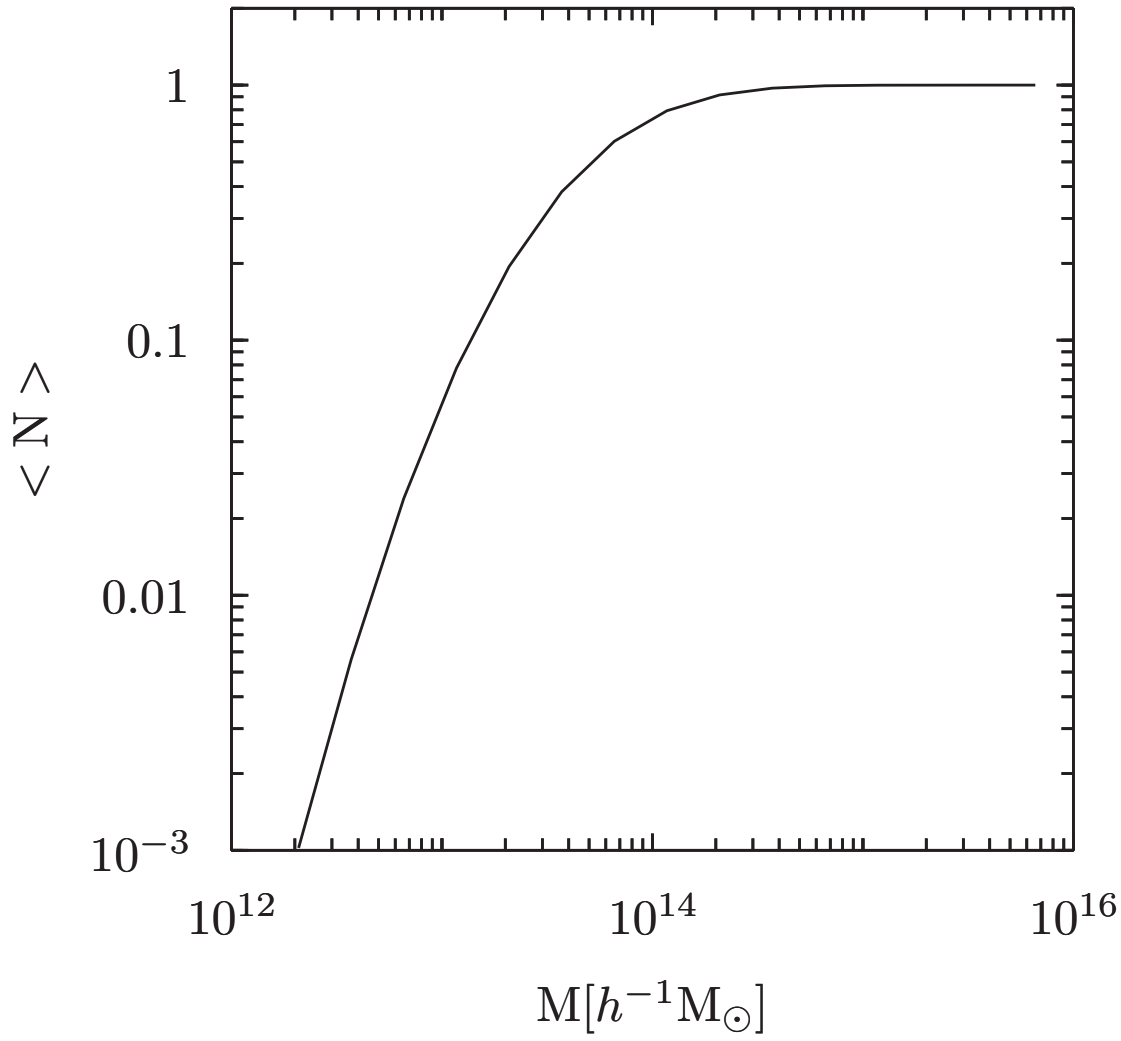


Figure 5.1: The average occupation number of central LRGs in host haloes of mass M, based on the model of Zheng et al. (2009).

0.07 $h^{-1}M_{\odot}$. In populating our haloes with LRGs we make the additional assumption that the above error bars in the model parameters – which are given to 1σ – can be modelled using a Gaussian distribution, which we use to introduce a similar error into our catalogue so as to mimic this uncertainty in the model. The halo occupation function for our haloes is shown in Figure 5.1. As can be seen from this plot there is a sharp drop-off in halo occupation below $10^{14}M_{\odot}$, to the extent that $10^{13}M_{\odot}$ haloes contain, on average, 0.05 LRGs.

5.2.2.2 Luminosity modelling

Now we have a population of LRGs in our haloes we need to assign properties to them, most importantly their luminosities. To do this, we rely solely on the mass of the host

haloes. The results presented in Zheng et al. (2009) indicate that the entire population of LRGs in their sample obeys the simple relation $L \propto \mathrm{M}^{0.66}$. Unfortunately, this is an inadequate prescription for assigning luminosities to our LRGs as, over the entire mass range of our host haloes, it results in too many unrealistically bright LRGs. A more detailed description of the L-M relationship is shown in Figure 3 of Zheng et al. (2009) which implies that at higher host halo masses the luminosity of LRGs does not scale as steeply as for lower masses. Zheng et al. (2009) discuss this result and make comparisons to other studies that shows a similar trend. For our modelling we adopt, based on their figure (Zheng et al., 2009), a relationship between mass and luminosity of the form:

$$L \propto \mathrm{M}^{\alpha}, \tag{5.11}$$

where the parameter α is given by

$$\alpha = \begin{cases} 1 & \text{for } \mathrm{M} \leq 5 \times 10^{11}\ h^{-1}\mathrm{M}_{\odot} \\ 0.5 & \text{for } 5 \times 10^{11}\ h^{-1}\mathrm{M}_{\odot} \leq \mathrm{M} < 5 \times 10^{12}\ h^{-1}\mathrm{M}_{\odot} \\ 0.3 & \text{for } \mathrm{M} \geq 5 \times 10^{12}\ h^{-1}\mathrm{M}_{\odot} \end{cases}. \tag{5.12}$$

This, combined with the comoving number density of LRGs in the sample, allows luminosities to be allocated to our LRGs in a manner that produces correctly the observed luminosity distribution of SDSS LRGs. We show a comparison of our model to the $M_g < 21.2$ SDSS sample in Figure 5.2. The SDSS data was based on the catalogue of Kazin et al. (2010), who closely match the previous catalogue of Eisenstein et al. (2005).

We apply this model to our data past the $z = 0.44$ limit of the modelling dataset. This is in order to create a base set of LRGs from all redshift slices in the simulation. For LRGs that exist at higher redshifts, this base dataset may require corrections. For example, various details of specific pipelines from observational catalogues can be readily incorporated onto this data, including any offsets from the LRG catalogue modelled here that may occur due to evolution effects.

5.2.2.3 Other LRG properties

The halo catalogue contains information on the locations of dark matter density peaks. The question of whether this corresponds to the locations of cluster BCGs has been recently studied by Zitrin et al. (2012), who used strong lensing to probe the underlying dark matter

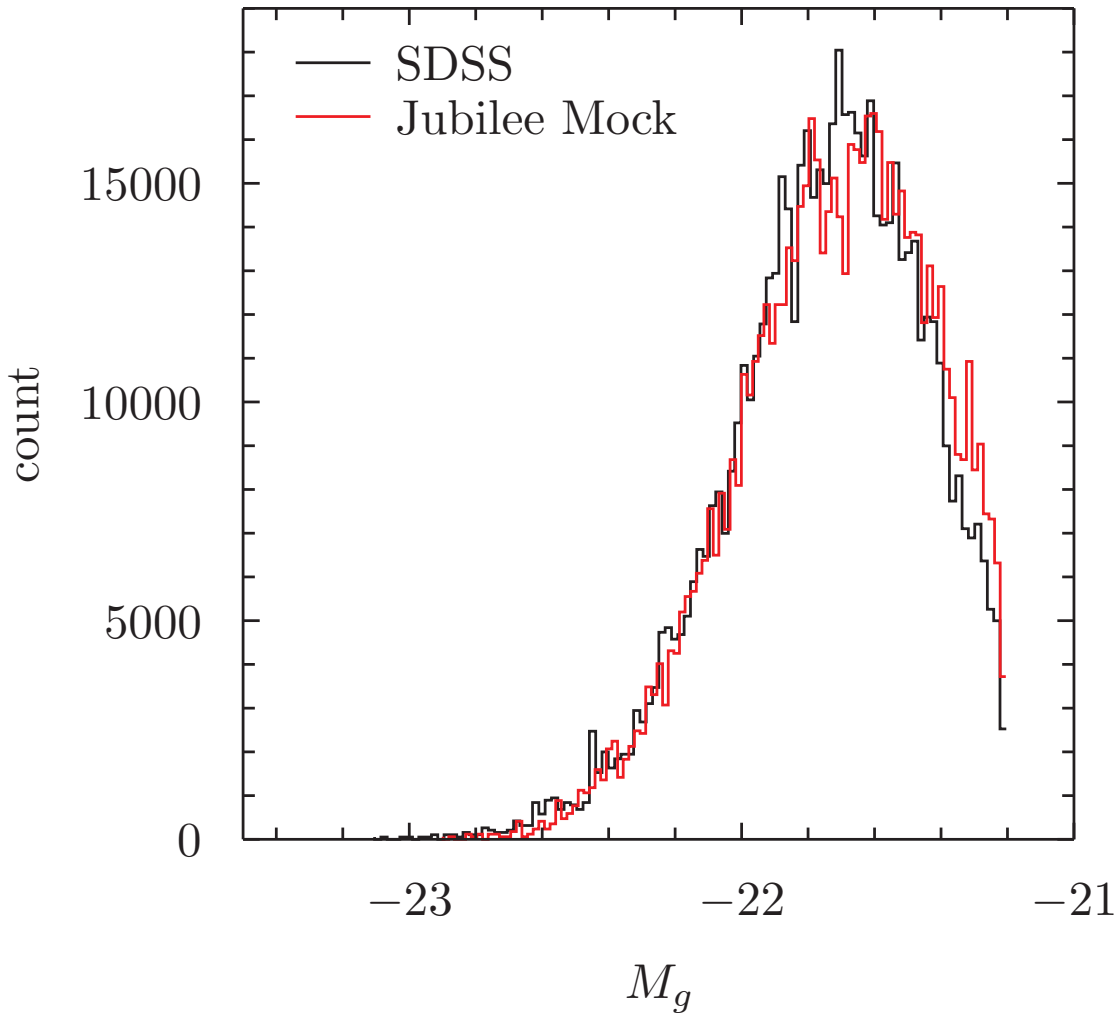


Figure 5.2: Histogram comparing SDSS LRGs with Jubilee mock LRGs. The full dataset of SDSS DR7 LRGs from Kazin et al. (2010) is plotted together with a random subsample of Jubilee Mock LRGs with the same total number count. The SDSS data is taken from a redshift range of $z = 0.16$ to 0.44 with a g-band absolute magnitude range of $M_g < -21.2$ (calibrated at $z = 0.3$). Jubilee mock data is taken from the $z = 0.3$ output slice.

distributions in 10,000 SDSS clusters. Their results show a small offset, with no preferred orientation, to the locations of BCGs from the dark matter density peaks. We apply their results to our dark matter halo catalogues in order to introduce this discrepancy between central LRGs, which we assume to be the BCGs in their particular haloes, and the underlying matter field that seeds the gravitational potential.

The results of Zitrin et al. (2012) showed that the scatter between the BCG location and density peaks are distributed log-normally in random directions via: $\log_{10}(\Delta_r\ [h^{-1}\mathrm{Mpc}]) = -1.895^{+0.003}_{-0.004}$. We produced a random scatter based on this and show the effect using a histogram in Figure 5.3. This figure should be directly compared to Figure 5 from Zitrin et al. (2012). We note that they observed what may be a trend with redshift to this scatter, that is, that the peak in Figure 5.3 would sit at ~ -2.5 for haloes at $z \sim 0.15$ and would evolve to ~ -1.7 for haloes at $z \sim 0.6$. However, the error bars in these results are large, being $\sigma \sim \pm 0.5$, and the trend of the evolution, if it exists, appears to be flattening out towards higher redshifts. Considering these facts, we do not attempt to parameterise the offset with redshift.

Finally, the velocities of the LRGs are assumed to be the same as the bulk velocities of their parent haloes. This assumption is likely to be slightly incorrect, as illustrated by Behroozi et al. (2013) who utilise a phase-space halofinder to show that halo cores frequently have an offset in velocity relative to the bulks of the parent haloes of $\sim 10 - 100\mathrm{kms}^{-1}$ depending on halo mass. In constructing sky maps of the LRGs it is possible to consider either a redshift that has been shifted due to the peculiar velocity of the LRGs or one that has not. In this chapter we consider LRGs with a Doppler-included redshift.

5.2.2.4 Simulated sky catalogues

With a complete set of LRGs in each of our output redshift slices it is possible to impose cuts to the catalogue in an attempt to mimic different observational catalogues. As an example, we model here the SDSS sample of Eisenstein et al. (2005), which is a natural choice because this is the sample on which Zheng et al. (2009) based their modelling. To create this catalogue we simply apply the magnitude cut from Eisenstein et al. (2005) onto our data. We assume that the catalogue covers the full sky and refer to it from this point as the 'SDSS mock' catalogue. In principle, other catalogues can be simulated by adopting constraints on magnitude and sky coverages, combined with subtleties such as completeness and scatter in photometric redshifts etc.

We show a histogram of LRG counts for both the SDSS mock catalogue and our entire

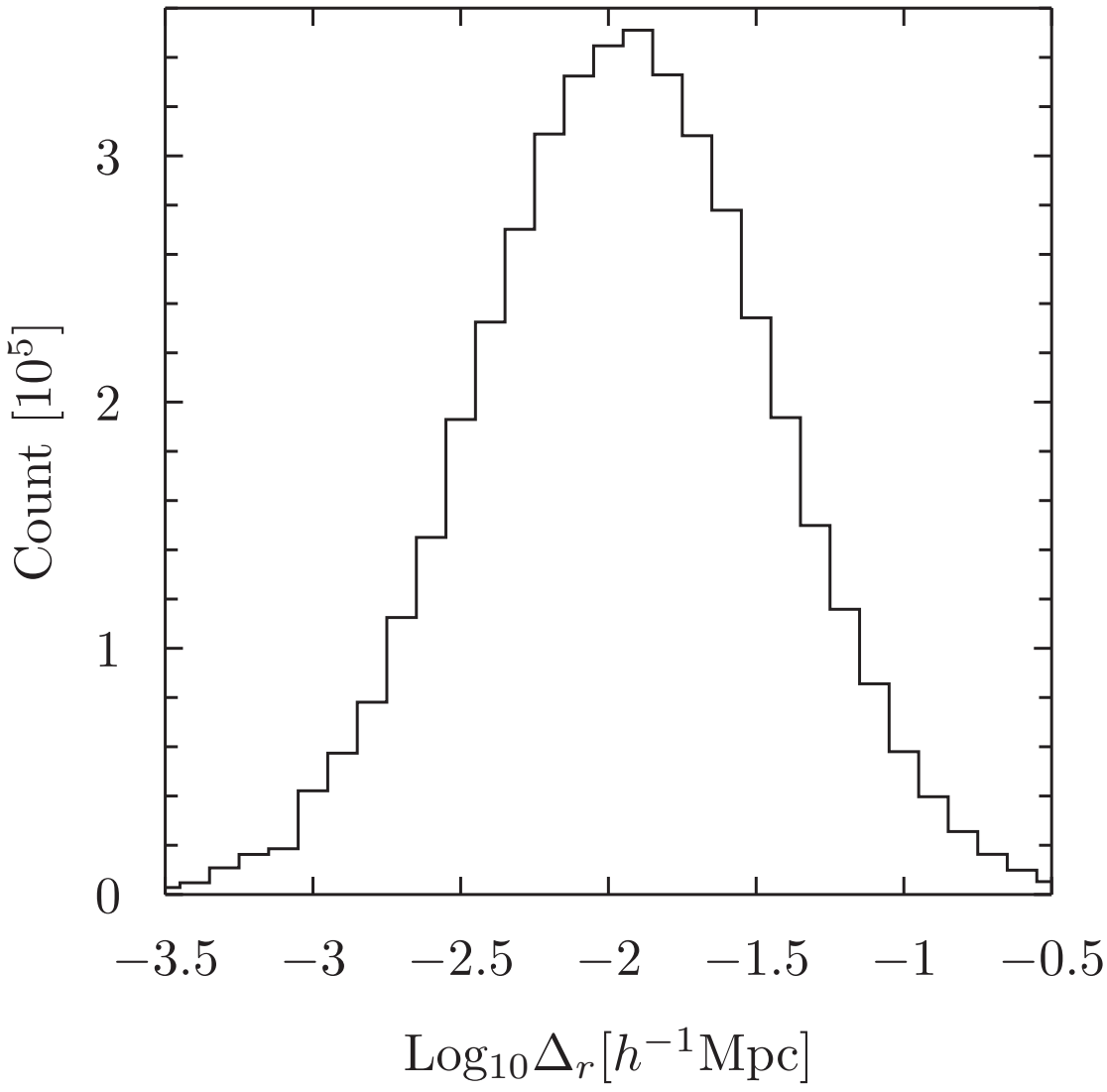


Figure 5.3: Distribution of random offsets between halo centres and the LRG locations, based on the results of Zitrin et al. (2012). The data shown here are based on 438,000 LRGs at $z = 0.5$.

sky catalogue of LRGs in Figure 5.4. The drop-off in LRG counts at low redshifts is due to the smaller volumes being sampled. The drop-off for $z > 1$ occurs because of the LRGs becoming rarer as the halo mass function evolves, cutting down the number of appropriately massive hosts as it does so.

5.2.2.5 Angular power spectra

Using a simulation places us in the privileged position of being able to sample galaxies across the entire sky with no masks in place. In this case, angular power spectra of the LRG catalogues are given by the standard relation given in equation 2.55, which we compute using the `map2alm` and `alm2cl` routines of HEALPIX. For the purposes of this chapter we consider full sky power spectra with no masks. When studying galaxy correlations in practice it is always necessary to consider corrections to equation 2.55 due to the use of masks that account for survey coverage and interference from undesired sources, particularly those in our galaxy. We correct the power spectra of our LRGs for shot noise using the relation

$$C'_\ell = C_\ell - \frac{4\pi}{N_{gal}}, \tag{5.13}$$

where $4\pi/N_{gal}$ is the expectation value of a random unclustered sample. The 4π term arises because we cover the full sky in steradians. Again, in practical surveys there is always less coverage than the complete sky and this term becomes some fraction of 4π.

5.2.3 Weak lensing maps

The weak lensing potential is identical to the ISW potential. That being so, we can calculate both the weak lensing and ISW effects from the same data. The lens equation is given by

$$\vec{\beta} = \vec{\theta} - \vec{\alpha}(\vec{\theta}, m(\vec{\theta})) \tag{5.14}$$

where $\vec{\alpha}(\vec{\theta})$ is the deflection angle created by the lens which depends on the observed positions, $\vec{\theta}$. We can write a dimensionless, integral version of equation 5.2 as

$$\Phi(r_p) = -\frac{G}{c^2}\int \frac{\rho(r_p - r'_p)}{|r_p - r'_p|} d^3 r'_p, \tag{5.15}$$

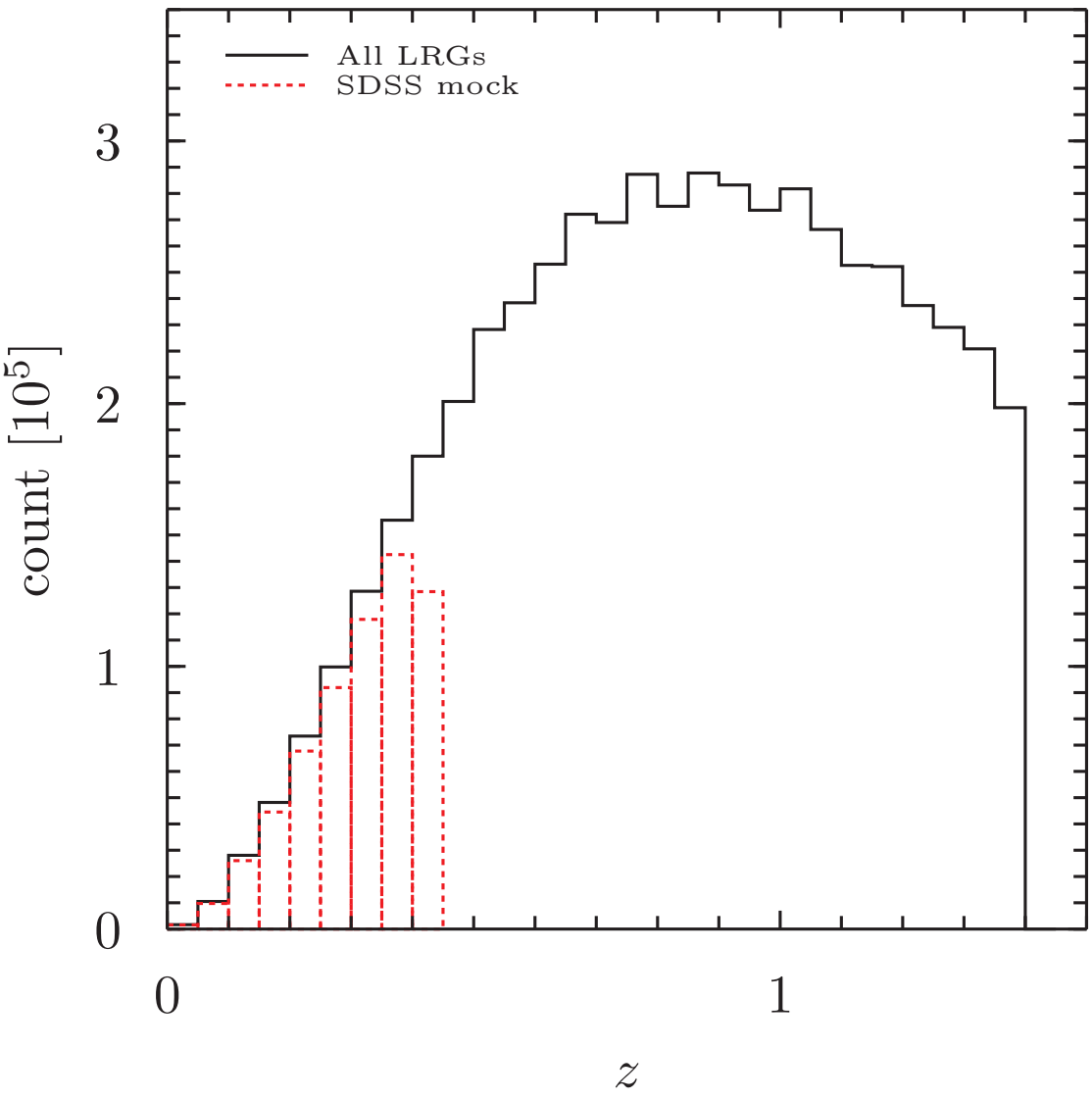


Figure 5.4: Histogram of LRG number counts by redshift. The count for our entire mock catalogue is shown in black. In red (dashed) we show the counts in our SDSS mock catalogue, which approximates the properties of the sample of Eisenstein et al. (2005) in the full sky.

where $r_p = (x, y, z)$ is a position in the box in physical units. Now we can define a new scalar (and adimensional) lensing potential in a given direction θ:

$$\psi(\theta) = \frac{2D_{ls}}{D_l D_s} \int \Phi(D_l\theta, z) dz, \tag{5.16}$$

where $r_p^2 = (D_l\theta)^2 + z^2$. The distances D_{ls}, D_l, and D_s are the angular distances from the lens to the source, the distance from the observer to the lens and the distance from the observer to the source, respectively. The relevant lensing quantities which interest us are then obtained from the derivatives of ψ. The derivatives are made with respect to the components of θ, i.e. (θ_1, θ_2). The deflection angle $\vec{\alpha} = (\alpha_1, \alpha_2)$ is given by the divergence of ψ and both the shear, $\vec{\gamma} = (\gamma_1, \gamma_2)$, and convergence, κ, are defined in terms of the second partial derivatives:

$$\begin{aligned}
\alpha_1(\vec{\theta}) &= \psi_1, & (5.17)\\
\alpha_2(\vec{\theta}) &= \psi_2, & (5.18)\\
\gamma_1(\vec{\theta}) &= \frac{1}{2}(\psi_{11} - \psi_{22}) = \gamma(\vec{\theta})\cos[2\varphi], & (5.19)\\
\gamma_2(\vec{\theta}) &= \psi_{12} = \psi_{21} = \gamma(\vec{\theta})\sin[2\varphi], & (5.20)
\end{aligned}$$

where $\gamma(\vec{\theta})$ is the amplitude of the shear, φ its orientation and

$$\psi_i = \frac{\partial\psi}{\partial\theta_i}, \tag{5.21}$$

$$\psi_{ij} = \frac{\partial^2\psi}{\partial\theta_i\partial\theta_j}. \tag{5.22}$$

The amplitude and orientation of the shear are given by

$$\gamma = \sqrt{\gamma_1^2 + \gamma_2^2}, \tag{5.23}$$

$$\varphi = \frac{1}{2}\text{atan}\left(\frac{\gamma_2}{\gamma_1}\right). \tag{5.24}$$

The convergence is

$$\kappa(\vec{\theta}) = \frac{1}{2}(\psi_{11} + \psi_{22}). \tag{5.25}$$

When measuring shear distortions, the reduced shear g is measured instead:

$$g = \frac{\gamma}{1-\kappa}. \tag{5.26}$$

Finally the magnification, μ, is:

$$\mu = \frac{1}{(1-\kappa)^2 - \gamma^2}. \tag{5.27}$$

All these quantities which describe the lensing effect can be obtained by combining the 5 derivatives, ψ_1, ψ_2, ψ_{11}, ψ_{22}, and $\psi_{12} = \psi_{21}$. For our particular case, given that we have simulation data in 3-dimensional Cartesian coordinates, it is convenient to express the derivatives of the lensing potential (originally with respect to the angle $\theta = (\theta_1, \theta_2)$) with respect to the physical coordinate $r = (x, y) = \theta D_l$. In these coordinates $\nabla_\theta = D_l \nabla_r$. The first and second derivatives with respect to θ of equation (5.16) can be rewritten in terms of derivatives with respect to $r = (x, y)$ as

$$\vec{\nabla}_\theta \psi(\theta) = F_{l1} \int \vec{\nabla}_r \Phi(x, y, z) dz, \tag{5.28}$$

$$\nabla^2_\theta \psi(\theta) = F_{l2} \int \nabla^2_r \Phi(x, y, z) dz, \tag{5.29}$$

where $F_{l1} = 2D_{ls}/D_s$ and $F_{l2} = 2D_l D_{ls}/D_s$. From our simulation outputs we convert the various Cartesian datasets into their sky projections by adopting the following coordinate system:

$$\begin{aligned} x &= \sin(\theta)\sin(\phi), \\ y &= \sin(\theta)\cos(\phi), \\ z &= \cos(\theta), \end{aligned} \tag{5.30}$$

where $\theta_1 = \theta$ and $\theta_2 = \phi$. We then compute the derivatives of Φ in the (θ_1, θ_2) coordinate system:

$$\begin{aligned}\Phi_1 = \frac{\partial \Phi}{\partial \theta} &= \frac{\partial \Phi}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial \Phi}{\partial y}\frac{\partial y}{\partial \theta} + \frac{\partial \Phi}{\partial z}\frac{\partial z}{\partial \theta}, \\ \Phi_2 = \frac{\partial \Phi}{\partial \phi} &= \frac{\partial \Phi}{\partial x}\frac{\partial x}{\partial \phi} + \frac{\partial \Phi}{\partial y}\frac{\partial y}{\partial \phi} + \frac{\partial \Phi}{\partial z}\frac{\partial z}{\partial \phi},\end{aligned} \tag{5.31}$$

$$\begin{aligned}\Phi_{11} = \frac{\partial \Phi_1}{\partial \theta} &= \frac{\partial \Phi_1}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial \Phi_1}{\partial y}\frac{\partial y}{\partial \theta} + \frac{\partial \Phi_1}{\partial z}\frac{\partial z}{\partial \theta}, \\ \Phi_{22} = \frac{\partial \Phi_2}{\partial \phi} &= \frac{\partial \Phi_2}{\partial x}\frac{\partial x}{\partial \phi} + \frac{\partial \Phi_2}{\partial y}\frac{\partial y}{\partial \phi} + \frac{\partial \Phi_2}{\partial z}\frac{\partial z}{\partial \phi}, \\ \Phi_{12} = \frac{\partial \Phi_1}{\partial \phi} &= \frac{\partial \Phi_1}{\partial x}\frac{\partial x}{\partial \phi} + \frac{\partial \Phi_1}{\partial y}\frac{\partial y}{\partial \phi} + \frac{\partial \Phi_1}{\partial z}\frac{\partial z}{\partial \phi}.\end{aligned} \tag{5.32}$$

From these derivatives we calculate the various lensing quantities on the sky as seen by an observer in the centre of the simulation volume using the relations

$$\begin{aligned}\psi_1 &= F_{l1}\Phi_1, \\ \psi_2 &= F_{l1}\Phi_2, \\ \psi_{11} &= F_{l2}\Phi_{11}, \\ \psi_{22} &= F_{l2}\Phi_{22}, \\ \psi_{12} &= F_{l2}\Phi_{12}.\end{aligned} \tag{5.33}$$

We assume in our analysis that the source object behind the lens is at a redshift of $z = 10$. All of our maps can be easily rescaled to simulation source objects that are at any redshift behind our lensing density fields, for example the CMB.

5.3 Results

5.3.1 ISW

In Figure 5.5 we show the projected dipole-subtracted ISW all-sky map from redshift $z = 0$ to 1.4. The negative blue regions correspond to projected under-dense regions where the dark-energy-driven acceleration of the expansion results in a net loss of energy for the CMB photons. On the other hand, when the CMB photons cross an over-dense region (red) the decaying potentials result in a net gain of energy for them. This map was constructed from a number of redshift shells and we show some of the maps from these

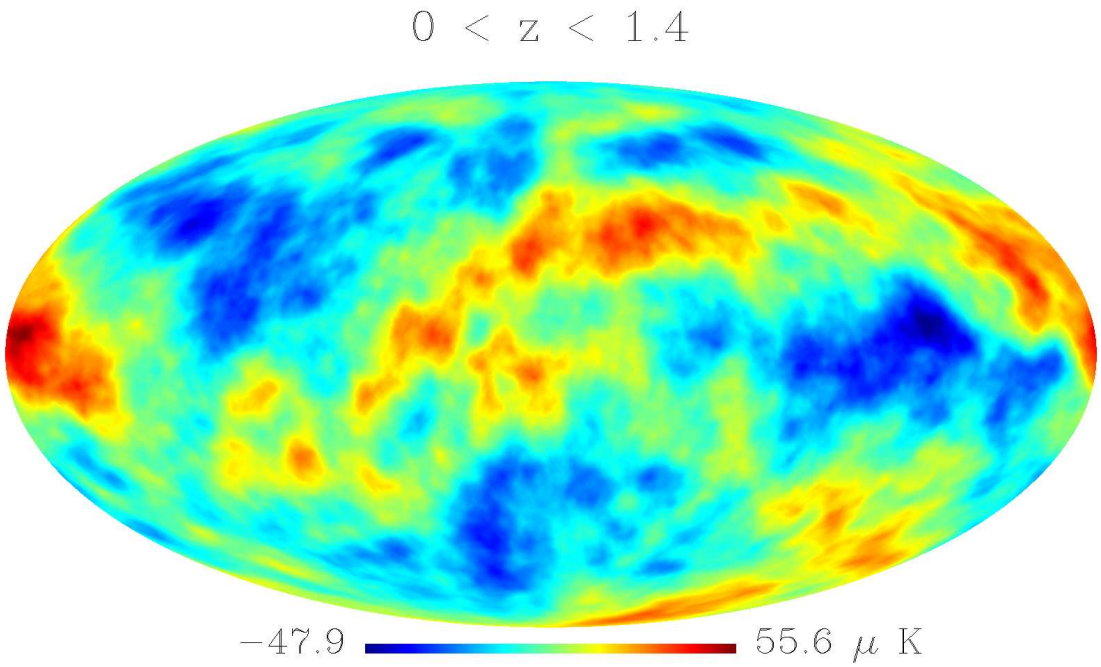


Figure 5.5: The full sky map of the predicted secondary CMB anisotropies due to the ISW effect from structures between redshifts of $z = 0$ to 1.4. The map is obtained by ray-tracing through the simulation potential field using the LAV approximation, as explained in § 5.2.1. The map is shown in Mollweide projection at a resolution of $N_{side} = 512$. The dipole contribution has been removed.

shells in Figure 5.6. This figure illustrates the varying imprint of the ISW anisotropies across redshifts. Lower redshifts show fluctuations over much larger areas of the sky than at higher redshifts, due to the changing angle that objects subtend in the sky at different redshifts. In addition, the amplitude of the anisotropies varies significantly across the redshift shells, as illustrated in Figure 5.7, where we show the redshift dependence of 1σ fluctuations in the anisotropies along with the temperature values of the hottest and coldest pixels for each redshift shell. The peak in this histogram is observed between $z = 0.2$ to 0.8, indicating that this is where the majority of the ISW contribution to the temperature anisotropies comes from. There still remains a significant amount of ISW effect generated past $z \sim 0.8$, however. The low-redshift drop in amplitude arises from the smaller volumes that are sampled at these redshifts. We explore in § 5.3.3 the implications of this plot on the ISW-LSS cross-correlation signal.

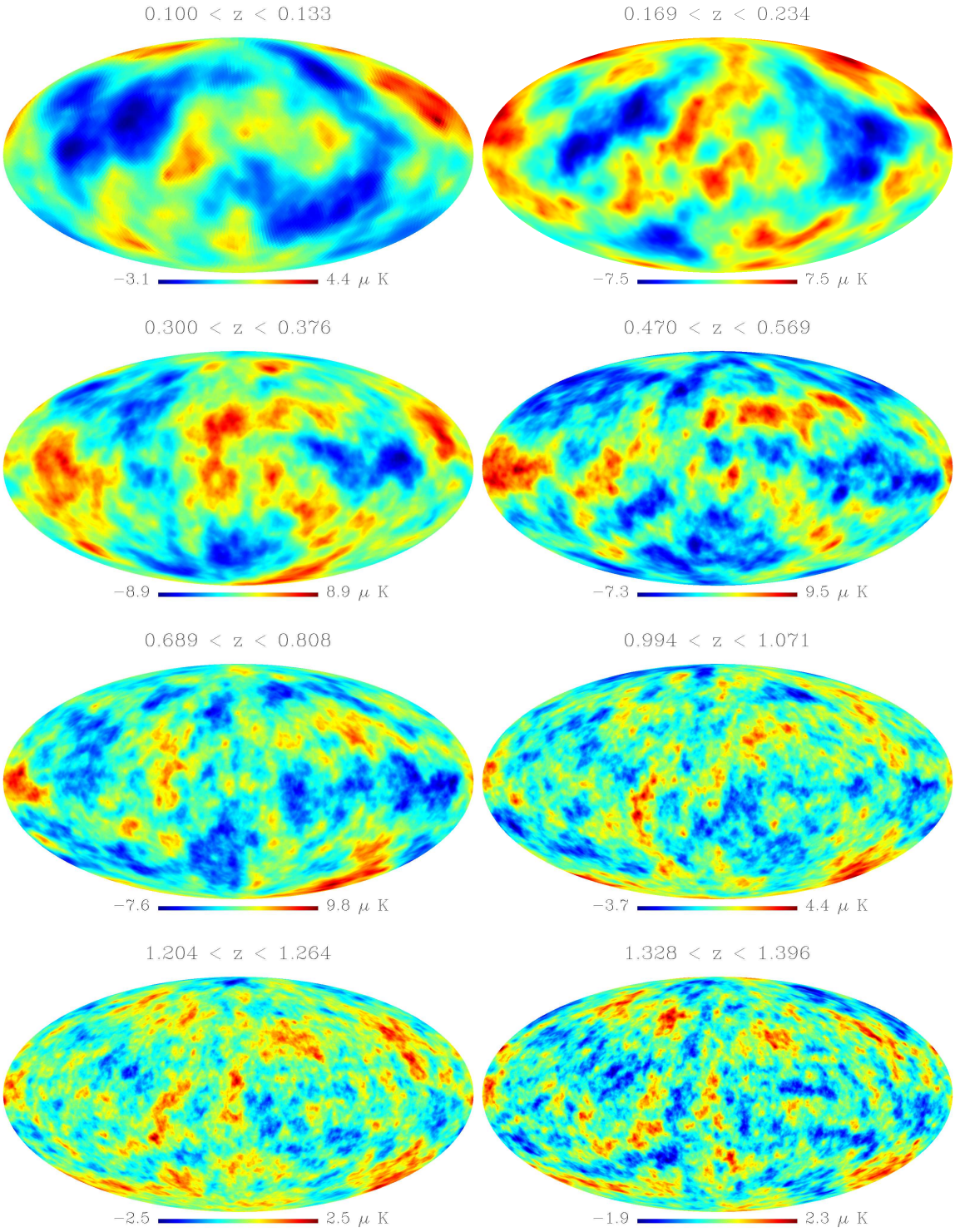


Figure 5.6: Full sky maps of the predicted secondary CMB anisotropies due to the ISW effect from structures between selected output redshifts. The maps are obtained by ray-tracing through the simulation potential field using the LAV approximation, as explained in § 5.2.1. The maps are shown in Mollweide projection with resolution N_{side} = 32, 128, 256 and 512 for redshifts of 0.100-0.133, 0.169-0.234, 0.320-0.569 and 0.689+, respectively. Dipoles have been removed from all maps.

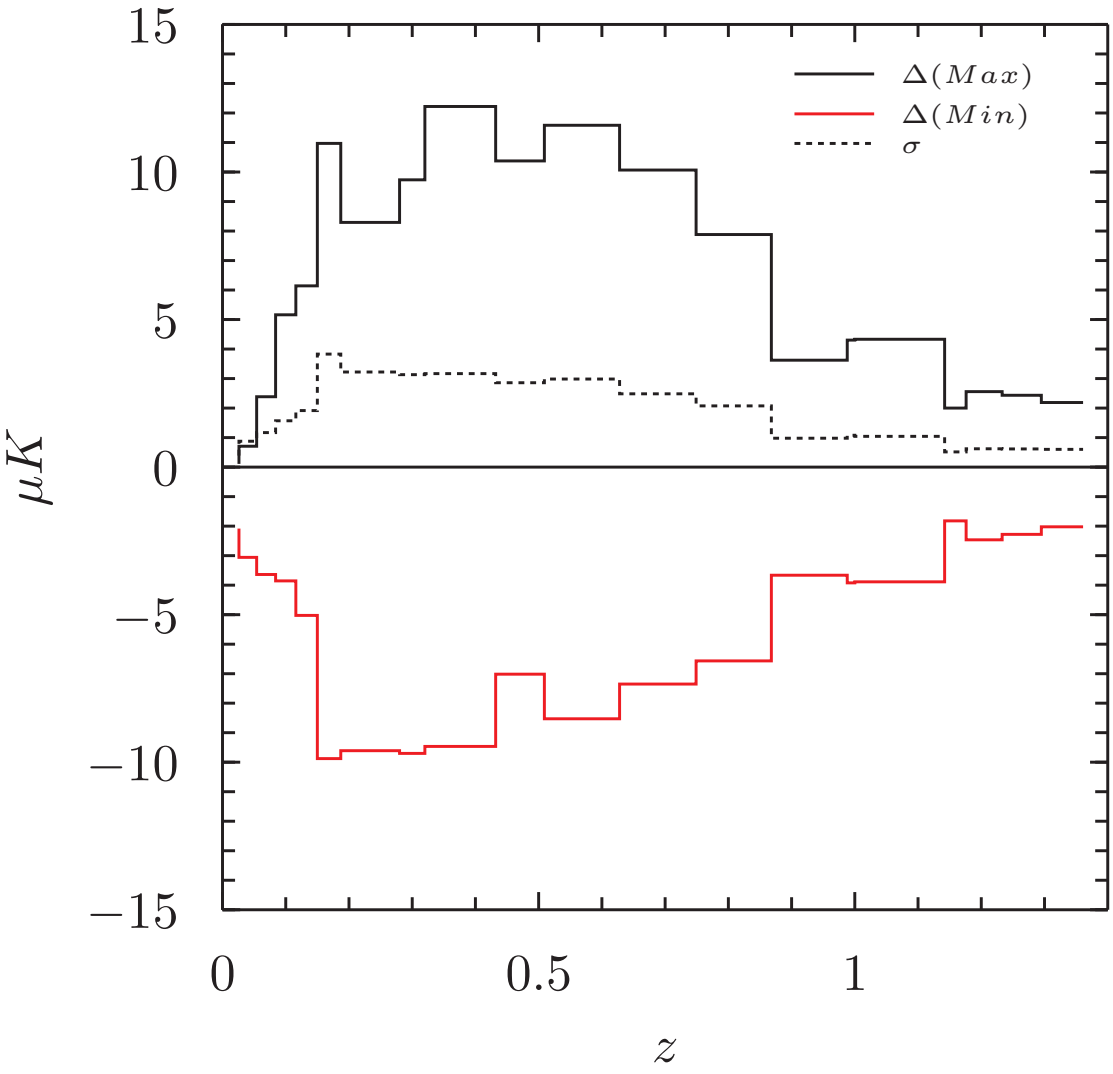


Figure 5.7: Hottest and coldest pixel values from the output maps of the ISW effect, from each of our 20 redshift bins. Also shown are the 1-σ values for the fluctuations in each redshift bin.

5.3.1.1 ISW power

The power spectrum of ISW-induced temperature anisotropies is shown in Figure 5.8. The spectrum shows a maximum at low-ℓ. At higher ℓs the slope of the spectrum follows a power law. This is the expected result using the LAV approximation. Cai et al. (2010) performed a detailed study of the contribution of the velocity field to the ISW effect, showing that the LAV power spectrum falls below that of the full ISW effect for higher values of ℓ. For example, the amplitude of the LAV ISW effect at $\ell \sim 100$ is around 50% of the full ISW amplitude and it drops down from $\sim 100\%$ at $\ell \leq 40$ (see Figure 17 of Cai et al., 2010). The under-representation of the ISW effect by the LAV approximation is redshift dependent with the drop-off from the full ISW effect, in general, occurring at lower ℓs for higher redshifts. Since our interest here lies in the dominant, low-ℓ part of the ISW effect, the LAV approximation is suitable for our purposes but the reader should be aware that results described for higher ℓs in this study are likely to slightly understate the reality of a full non-linear ISW effect. We discuss in § 5.4.4 the impact that taking the LAV approximation has on the cross-correlation results.

In Figure 5.8 we show the power spectra of the ISW effect anisotropies in different redshift bins (0−1.4, 0−0.4, 0.4−0.8, 0.8−1.2) and compare them to predictions from linear theory. The figure shows that in our simulation, due to our large box size, we are able to view the ISW effect on very large-scales without an appreciable drop-off in power. This illustrates the requirement, when simulating the ISW, for a box that captures very large-scale fluctuations in the density field. For the ℓ-range under consideration here the results from the simulation correspond closely to those from linear theory, as per the study of Cai et al. (2010), who found that the LAV matches linear theory to well past $\ell \sim 100$. The power spectra in Figure 5.8 have been binned. The low-ℓ data points ($\ell < 6$) are taken in bins of width $\Delta\ell = 1$ and show scatter from cosmic variance. Because we model a volume with a side-length of 6 h^{-1}Gpc, we capture much of the large-scale power in the potential. Despite this, the low-ℓ regime of Figure 5.8 shows that we may be losing a small amount of power on these scales, although not nearly to the extent of that observed in the tiled 1 h^{-1}Gpc box used in Cai et al. (2010).

5.3.2 LRGs

In Figure 5.9 we show a sky map of LRG number counts from all the LRGs in our catalogue between $z = 0$ to 1.4. No cuts of any kind have been applied to this figure thus it represents the spatial positions on the sky of all the LRGs underneath the black, solid line

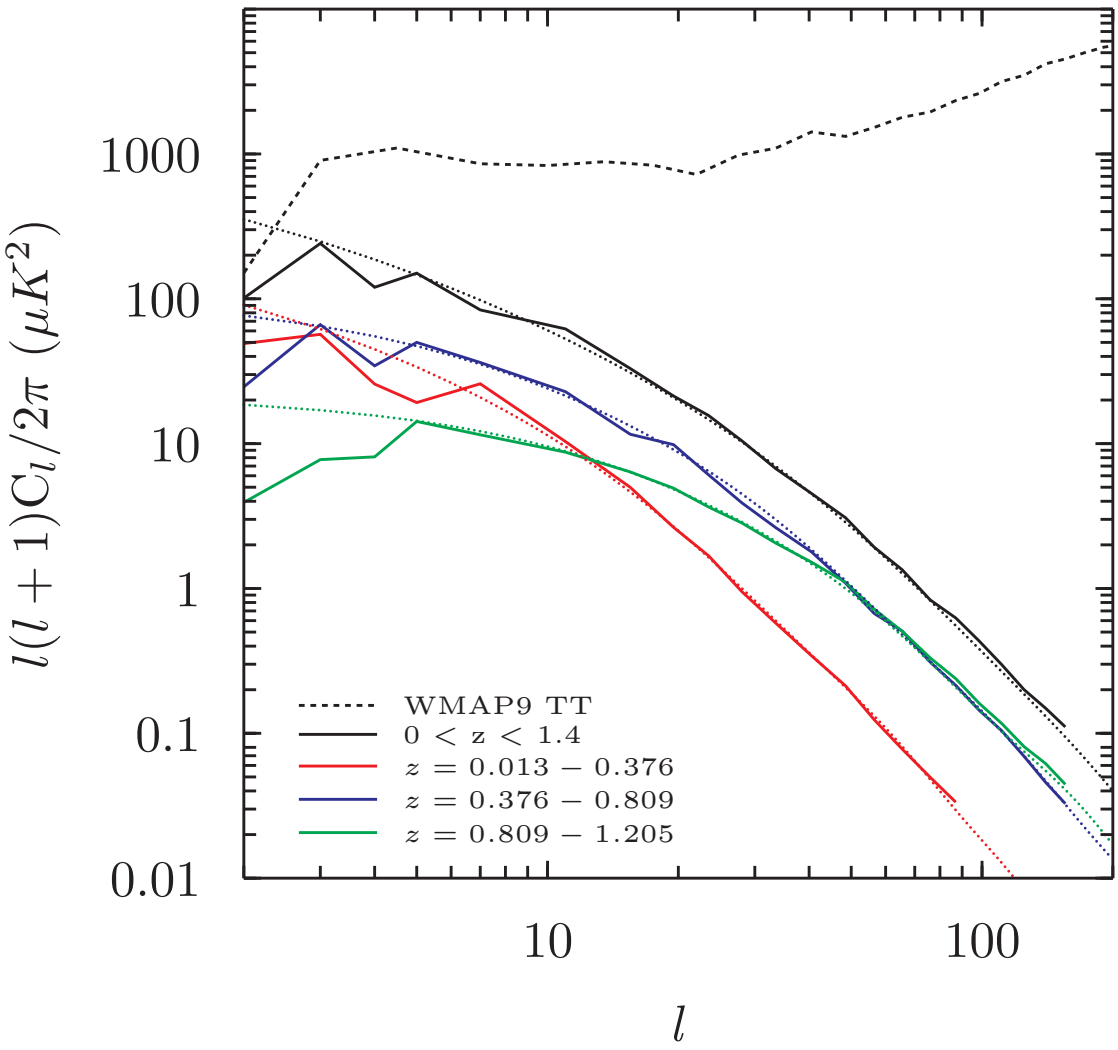


Figure 5.8: The power spectrum of the temperature anisotropies that arise from the ISW effect. Power from individual redshift bins are shown along with the full integrated ISW effect (from $z = 0$ to 1.4) and the CMB TT power spectra. Linear theory predictions are shown as dotted lines. The redshift bin widths correspond to those of Figure 5.7.

in Figure 5.4. Figure 5.10 shows a projection of the LRGs in the simulation by distance from the observer. Both panels represent a projection that is 20 h^{-1}Mpc deep, with the top panel showing all LRGs out to a radius of 3 h^{-1}Gpc ($z \leq 1.4$) and the bottom panel showing a zoomed-in view of the LRGs out to a radius of 500 h^{-1}Mpc ($z \leq 0.17$). Voids and filamentary structures are clearly seen in the distribution. There is little distortion from the peculiar motions of the LRGs. This is because the LRGs are all central galaxies and have been assigned the bulk velocity of their host haloes. Therefore their peculiar velocities are small compared to the higher peculiar velocities of satellite galaxies which orbit the centre of mass of a cluster and create the distinctive 'Fingers-of-God' effect.

We show the angular power spectrum of our simulated, full sky catalogues in Figure 5.11. The data has been split into the same redshift shells that we show in Figure 5.7. The results in Figure 5.11 show the expected trend that, as structure formation proceeds, correlations between galaxies grow stronger. We also plot the ISW effect power spectrum on Figure 5.11 alongside the LRG power which reveals that the interplay between the LSS and ISW spectra makes it difficult to observe a strong cross-correlation: where the LRG power is high the ISW power is low and vice versa. At both high and low-ℓs, therefore, the expected cross-correlation should fall off, leaving only the region around $\ell \sim 10 - 40$ where we should expect a strong cross-correlation.

5.3.3 ISW correlation with LSS

For our cross-correlation analysis we now investigate how redshift selection of LRGs affects the strength of the ISW-LSS correlation signal. The results and discussion presented here relate to the signal-space for measurements of the ISW-LSS cross-correlation. We stress that this is different from detection-space, in that no signal-to-noise considerations are included in this analysis. In Figure 5.12 we calculate the cross-correlation signal between the ISW effect from $z = 0$ to 1.4 and LRGs using the same redshift shells as in Figure 5.11. The results show that covering the peak of the contribution to the ISW effect, in terms of redshift ($z \sim 0.2$ to 0.8), is an important factor in producing the signal. The lower redshift bins appear to show a peak in the signal at lower values of ℓ, although this is difficult to judge because there is significant scatter in the data due to cosmic variance in this regime. The ISW-SDSS cross-correlation, shown in Figure 5.13, shows a similar trend to the lower redshift ISW-LRG signals, as expected.

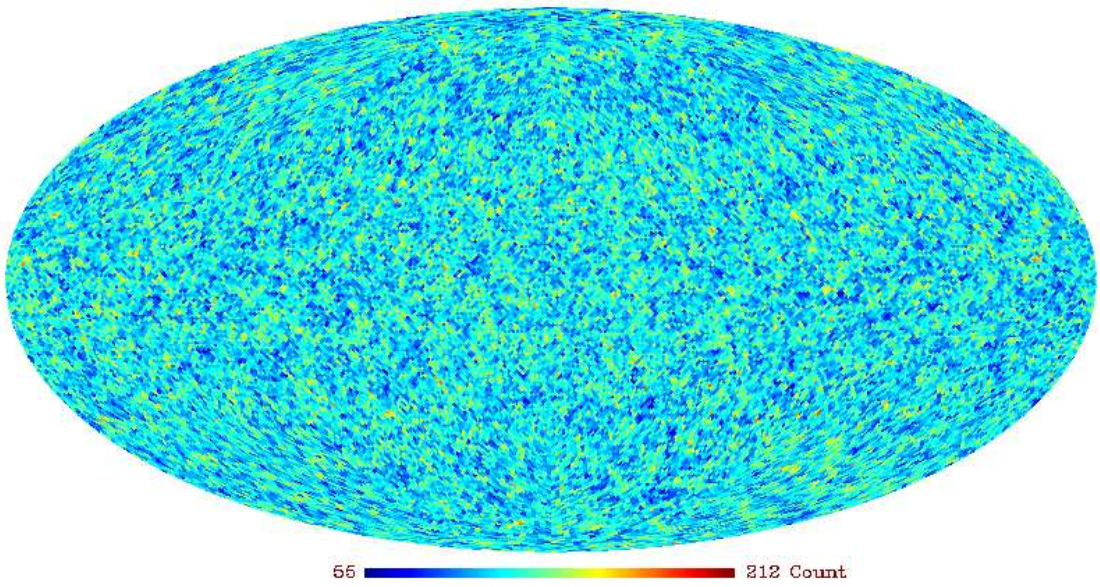


Figure 5.9: All-sky projection of the Jubilee mock catalogue LRGs from $z = 0$ to 1.4. A Mollweide projection has been used with $N_{side} = 64$. No cuts have been applied to the data.

5.3.4 Lensing maps

We have produced all-sky lensing maps of the various quantities mentioned in § 5.2.3, in particular the convergence, κ, and the deflection angles α_1 and α_2. We show in Figure 5.14 a collection of complementary plots for the $z = 0.150$ redshift shell, which spans a redshift range of $z = 0.13$ to 0.17. The plots show the projected density field (Figure 5.14a), which can be seen closely matches the convergence (Figure 5.14b), as expected. We also show the ISW map (Figure 5.14c) and the effect of a very large overdensity in the right-centre of the plot is very clear, creating as it does a deep potential well and a strong ISW-induced temperature anisotropy. The α_1 and α_2 plots (Figure 5.14e and 5.14f) in the region of this large overdensity show characteristic dipoles in the orthogonal θ and ϕ directions (which for this cluster, as it is on the equator of the plot, can be thought of as roughly the same as the up-down and left-right directions respectively). The amplitude of the combined α_1 and α_2 deflection angles (Figure 5.14d) shows very clearly the large overdensity affecting light rays in its region of the sky and features in this map can be seen to correspond to ones in the ISW map.

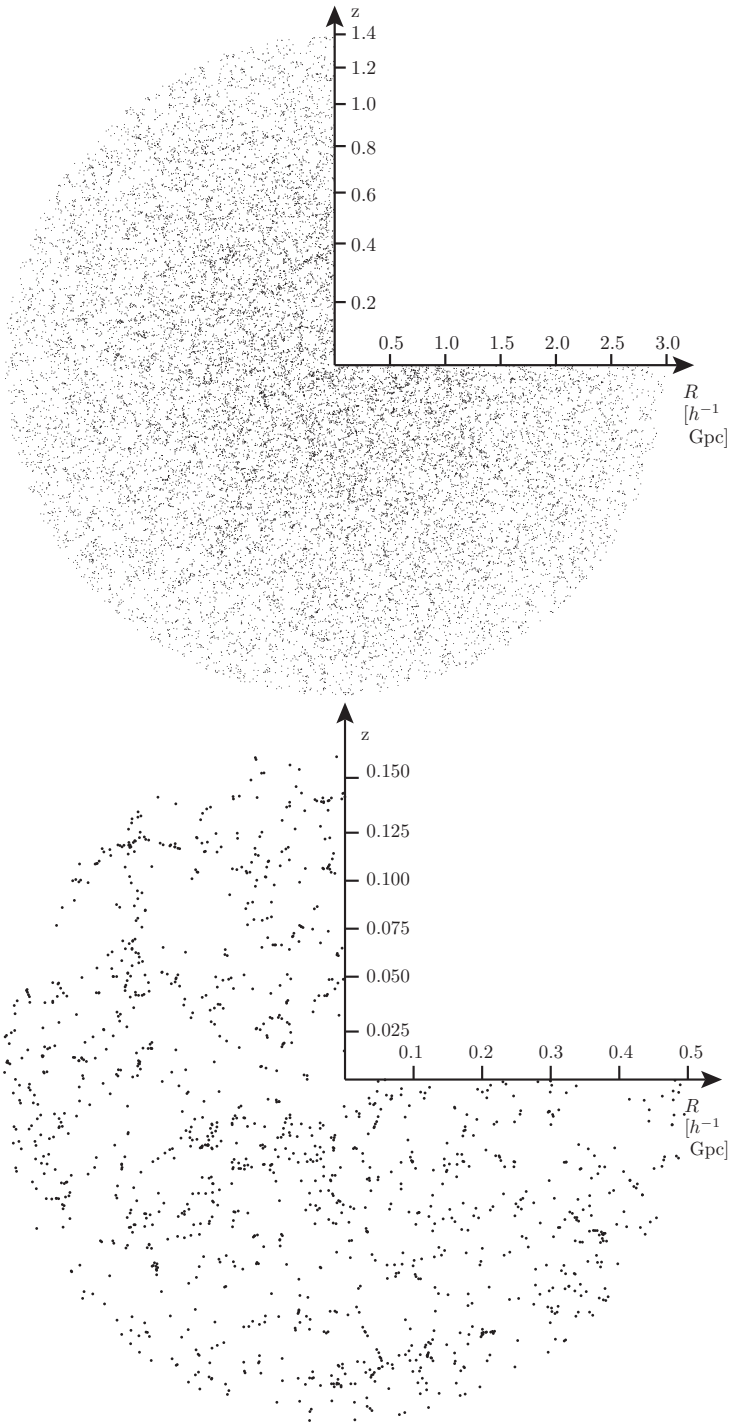


Figure 5.10: Projections of the full LRG catalogue distribution from $z = 0$ to 1.4. *Top panel:* All LRGs lying in a 20 h^{-1}Mpc thick slice, within 3 h^{-1}Gpc of the observer in the centre of the box. *Bottom panel:* A zoom-in of the local LRG distribution between $0 - 0.5$ h^{-1}Gpc from the observer.

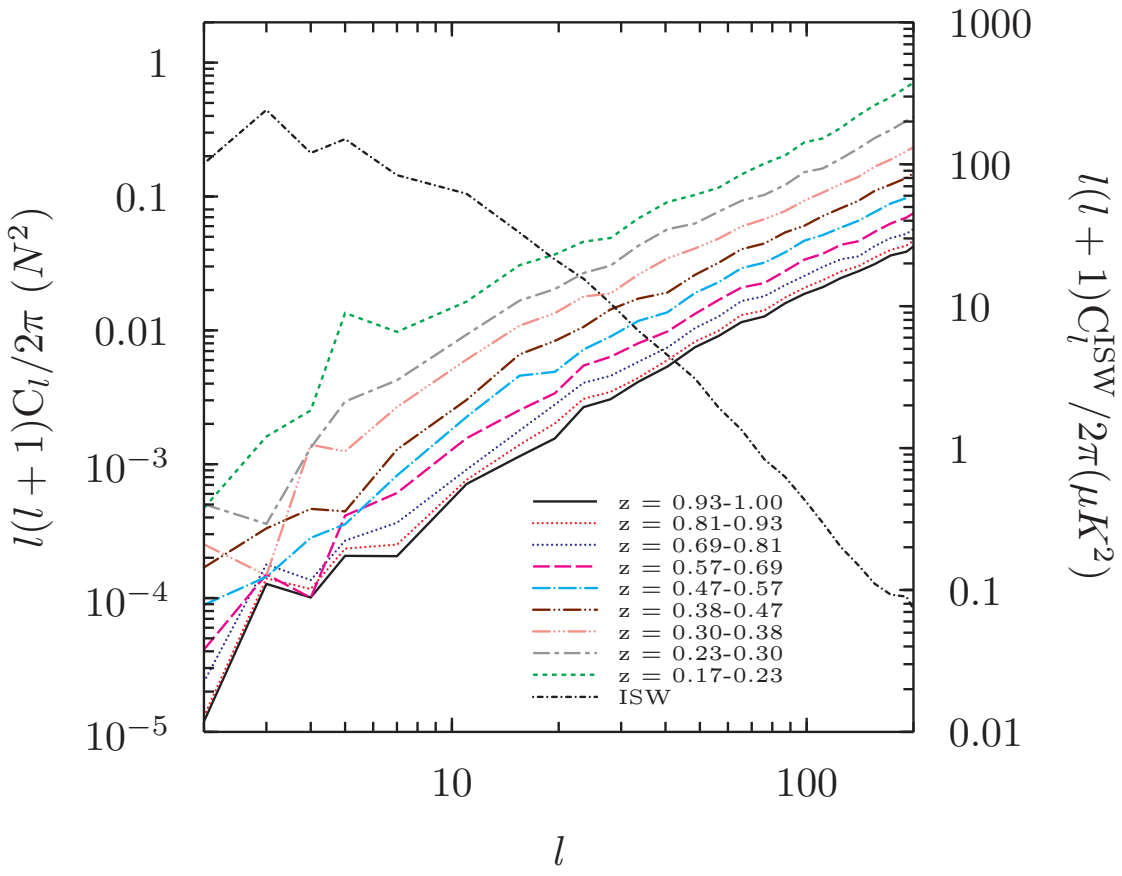


Figure 5.11: Power spectra of LRGs for the redshift bins shown in Figure 5.7. Overlaid using the second y-axis we also show the ISW power spectrum from Figure 5.8.

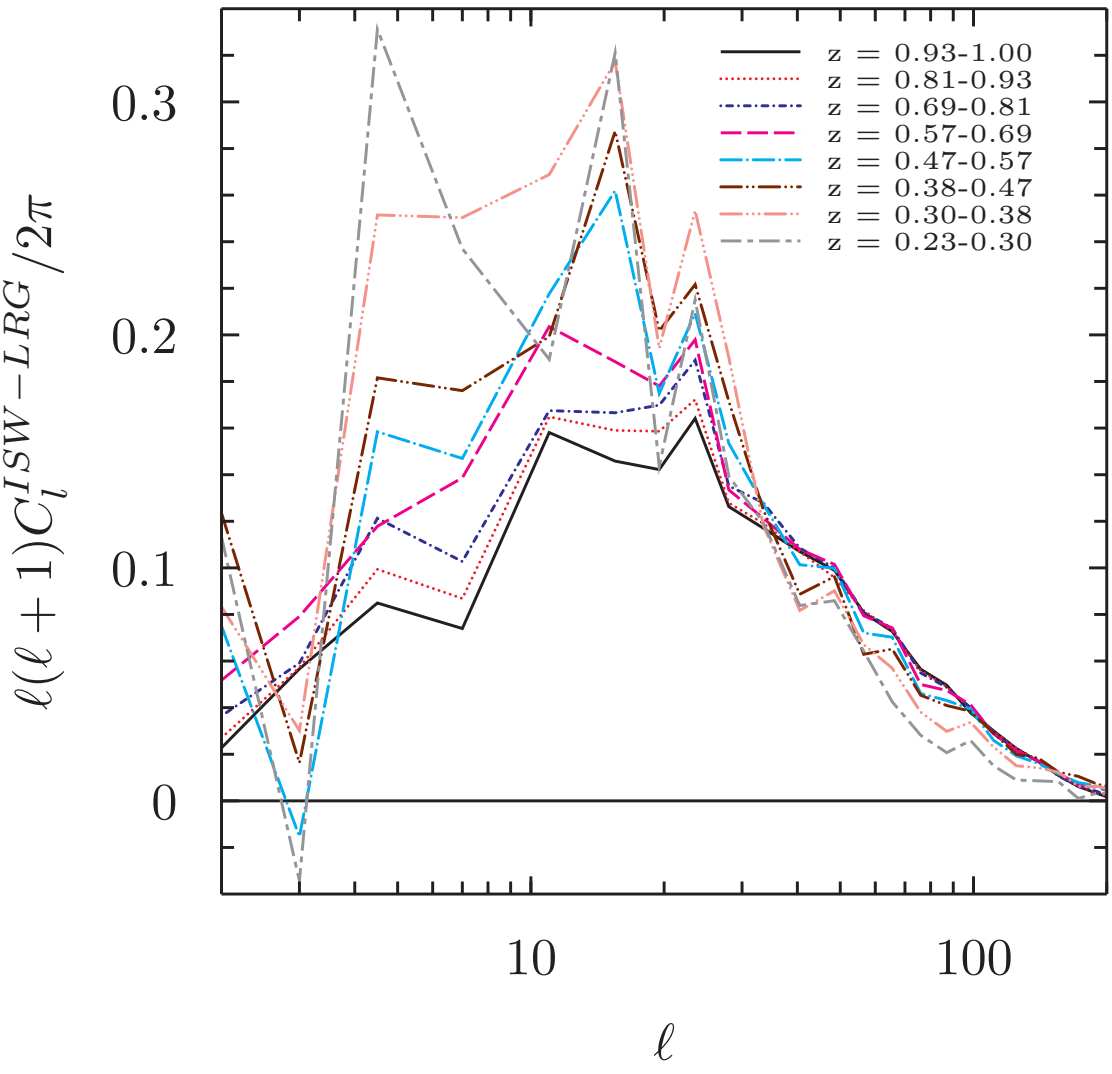


Figure 5.12: Cross-correlation signal between the LAV approximation ISW effect (calculated from $z = 0$ to 1.4) and all LRGs split into the same redshift bins as Figure 5.11.

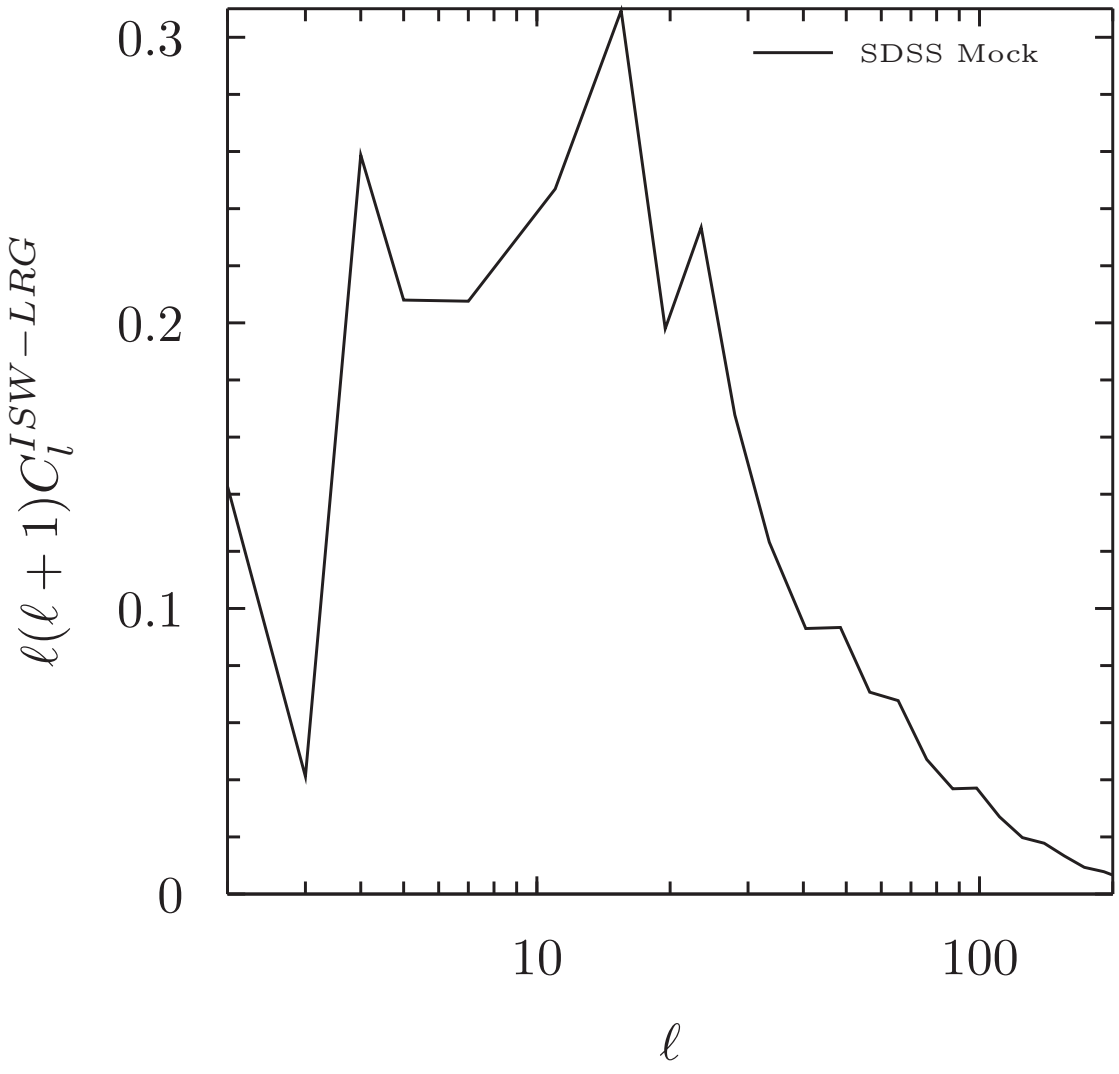
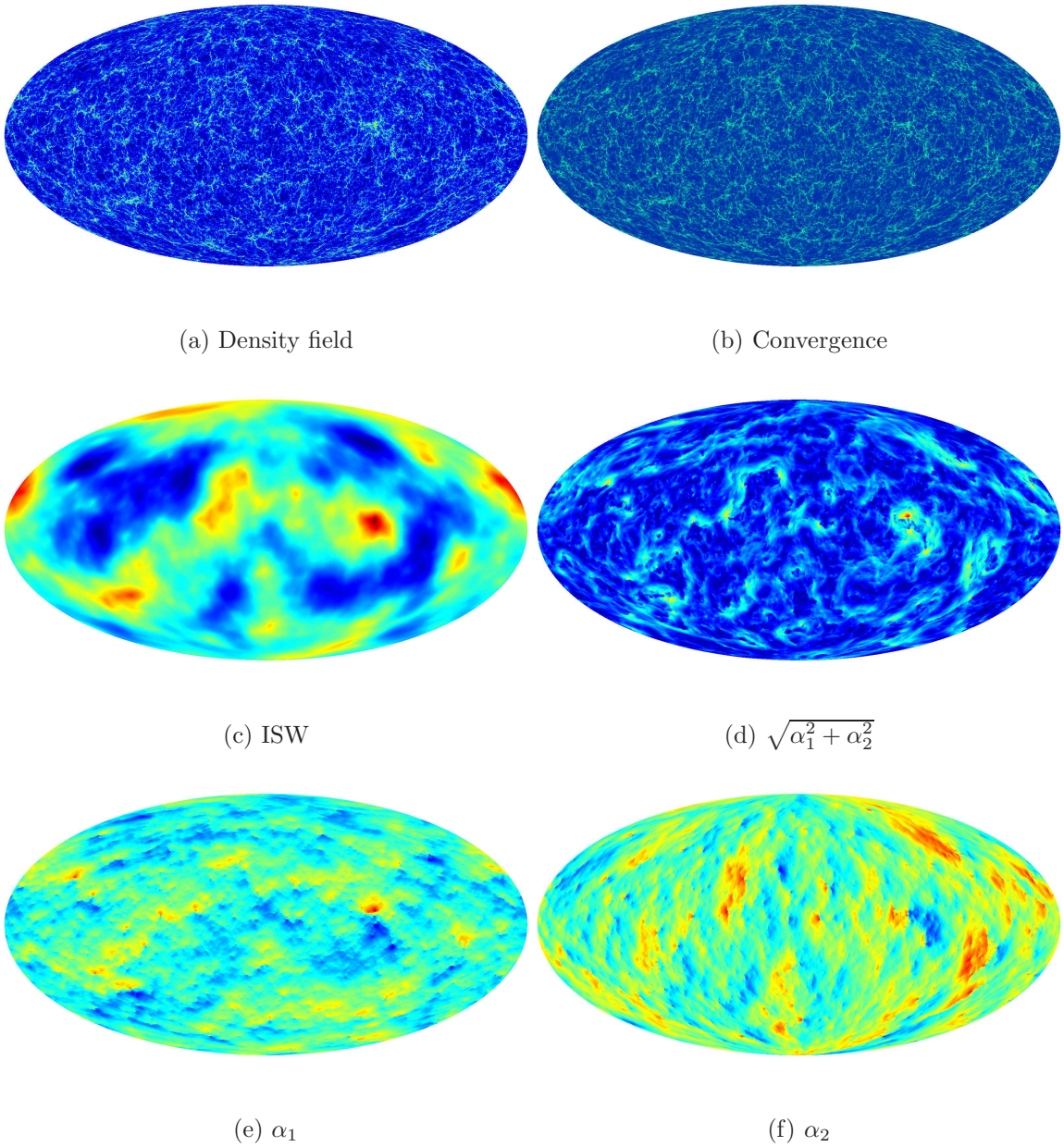


Figure 5.13: Cross-correlation signal between the LAV approximation ISW effect between $z = 0$ to 1.4 and LRGs from the mock SDSS survey.



(a) Density field

(b) Convergence

(c) ISW

(d) $\sqrt{\alpha_1^2 + \alpha_2^2}$

(e) α_1

(f) α_2

Figure 5.14: Complementary lensing and ISW maps from the $z = 0.150$ redshift shell, which spans a redshift range of $z = 0.13$ to 0.17.

5.3.5 Online databases

All our LRG data will be publicly available online at http://jubilee-project.org. An SQL database has been set up so that the data can be queried to suit the requirements of individual users. In addition to various mock LRG surveys we will also be providing halo catalogues, void catalogues, an NVSS-like radio catalogue, as well as sky maps including lensing maps and density fields.

5.4 Discussion

5.4.1 Implications for survey characteristics

Figure 5.7 illustrates a general point regarding how the ISW signal can be measured observationally. A cross-correlation approach attempts to use as large a galaxy survey as possible in order to uncover as much of the information in Figure 5.7 as is possible. An appropriate LSS survey must therefore cover a redshift range that complements the range seen in Figure 5.7. This is illustrated in Figure 5.12 where we see that the signal is strongest for redshift shells that cover the peak in the ISW effect at $z \sim 0.2 - 0.8$. This is not necessarily a result that is applicable in practice because in ISW detection efforts the signal-to-noise of the particular observational strategy needs to be considered. For example, sky coverage is a major stumbling block in the LSS cross-correlation approach because anything less than a full sky survey begins to impact the signal-to-noise level of the detection. This places a strain on any LSS survey – which typically have to balance sky coverage versus survey depth – that aims to optimise an ISW measurement. A detailed analysis of the various signal-to-noise considerations in the LSS-ISW correlation measurement can be found in Cabré et al. (2007).

Previous correlation measurements have utilised a variety of LSS catalogues including the NVSS (Condon et al., 1998) radio survey, which had a sky coverage of 82%, and the SDSS galaxy survey with a sky coverage of 35% (Ahn et al., 2012). These surveys have their own pros and cons. Whilst the NVSS has excellent sky coverage it has only ~ 1.4 million objects and these are found across a wide redshift distribution ($z = 0$ to 2+, we refer the reader to Figure 2 from Planck Collaboration, 2013b). The SDSS, on the other hand, contains many sources (almost 1 billion galaxies in total), in particular LRGs, across a redshift range ($z \sim 0$ to 0.8, see Figure 2 of Planck Collaboration, 2013b) that is very well suited to ISW effect detection. However, because of its smaller sky coverage, it has a high noise level on larger scales.

There are future surveys that will have appropriate sky and redshift footprints, in particular the HI Evolutionary Map of the Universe (EMU) survey (Norris, 2011), that will be performed using the ASKAP telescope. EMU is a pathfinder for the Square Kilometre Array (SKA), and will detect sources across a broad range of redshifts, $z \sim 0-6$, in particular low-redshift star-forming galaxies at $z < 2$. Its sky coverage will be roughly the same as for the NVSS and the intention is for its data to be combined with another HI survey, Westerbork Observations of the Deep APERTIF Northern sky (WODAN) (Röttgering et al., 2011), which will cover the remaining patch of the Northern Hemisphere that EMU cannot see.

5.4.2 Stacking and ΛCDM tension

Recent studies of stacked CMB patches (Nadathur et al., 2012; Hernandez-Monteagudo & Smith, 2012; Ilic et al., 2013) have focussed on corroborating the result of Granett, Neyrinck & Szapudi (2008). This result, based on the average signal from 50 overdense and 50 underdense lines of site found from SDSS data, was: $\langle \Delta T \rangle = +7.9 \pm 3.1\ \mu K$ and $\langle \Delta T \rangle = -11.3 \pm 3.1\ \mu K$, for overdense and underdense regions respectively. Taken together the data suggested a mean temperature deviation of $\langle \Delta T \rangle = +9.6 \pm 2.2\ \mu K$, which constitutes a detection of the ISW effect with a significance of over 4σ.

In their methodology Granett, Neyrinck & Szapudi (2008) filtered the CMB signal found behind each of their patches with a compensated top-hat filter of radius 4°, corresponding to physical scales of 100 h^{-1}Mpc. The filter took the form:

$$W(\theta) = \begin{cases} 1, & 0 \leq \theta \leq \theta_c, \\ -1, & \theta_c \leq \theta \leq \sqrt{2}\theta_c, \end{cases} \tag{5.34}$$

with $\theta_c = 4^{\circ}$. This filter had the effect of nullifying the influence of CMB fluctuations on scales larger than the aperture. One immediate consequence of this is that the temperature anisotropies measured with this filter in place are not comparable to, for example, the values shown in Figure 5.7. However, we can make a comparison to the results of Granett, Neyrinck & Szapudi (2008) by applying the same filter onto the full sky ISW map of Figure 5.5. The result of this filtering is shown in Figure 5.15, which has absolute maximum and minimum fluctuation values of $+7.3\ \mu K$ and $-8.4\ \mu K$ respectively. On the face of it, these values do not seem too incongruous with the result of Granett, Neyrinck & Szapudi (2008), however, if we were to consider a patch of sky that covers 35% (as per the SDSS data) and then consider the average value of 50 overdense and 50 under-

dense regions in that patch, the actual measured fluctuation value will drop considerably. Analysis by Nadathur et al. (2012) has shown that the result of Granett, Neyrinck & Szapudi (2008) is inconsistent with linear theory predictions of ΛCDM at a level of $> 3\sigma$. Hernandez-Monteagudo & Smith (2012) have confirmed the magnitude of Granett et al.'s original result and have also shown that appealing to higher-than-linear aspects of the ISW mechanism, i.e. the Rees-Sciama effect, does not account for the extra signal (not least because a smoothing filter of $3-4^\circ$ has been applied in most stacking analysis, thus averaging over non-linear contributions to the signal that have been shown by Cai, Cole, Jenkins & Frenk (2010) to begin to dominate the signal on scales less than 1°).

This implies that either ΛCDM might not be correct or that there are unaccounted for systematics in the stacking process. The latter possibility might be the case as noted by Ilic et al. (2013). One possibility is selection bias in the choice of lines of site. It is possible that the voids or superstructures considered typically overlap with others, for example. This is unlikely to be the case for all 50 lines of site, however. In terms of systematic effects occurring in the stacking procedure in general it should be noted that the appeal of the approach is that it connects two unrelated datasets in the CMB and LSS. The selection of underdense or overdense regions should be totally independent of the values of the CMB temperature fluctuations that exist behind them (aside from ISW-induced effects of course).

The findings of Planck Collaboration (2013b) using stacking techniques was a detection of the ISW signal with a significance of $\sigma \approx 0-4$. The range of significances depended on which catalogue they used, with $\sim 4\sigma$, the strongest result, applied to the same catalogue as Granett, Neyrinck & Szapudi (2008), showing 'strong discrepancies in amplitude and scale compared to expectations'. The range of significances found using more recent catalogues was $\sigma \sim 0-2.5$ with a 'more consistent scale and amplitude'. These results are suggestive that the stacking approach still needs to be thoroughly understood in terms of systematics.

5.4.3 Model discrimination using the ISW

The hope that the ISW signal in the future can be used to help discriminate between cosmological models depends on our ability to measure the signal accurately. Results from stacking approaches, discussed in the introduction to this chapter, are currently placing the ΛCDM model under scrutiny. Work by previous authors on cross-correlations between the CMB and the ISW have attempted to constrain cosmological parameters (Padmanabhan

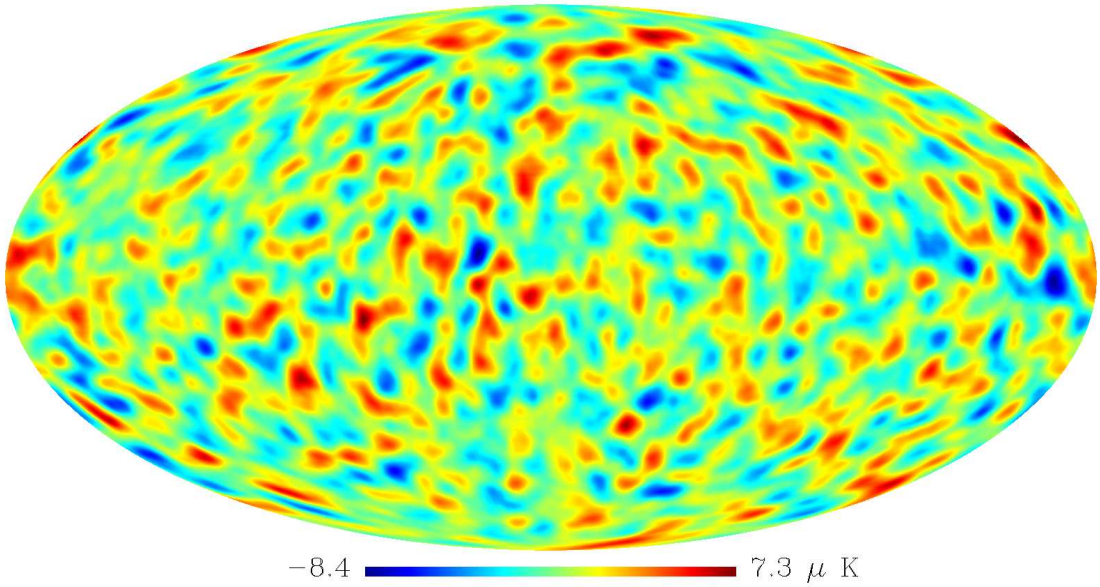


Figure 5.15: The all sky ISW map filtered with a compensated top-hat filter of radius 4°. The dipole was removed prior to filtering.

et al., 2005; Pietrobon et al., 2006; Gaztañaga et al., 2006; Giannantonio et al., 2008). These results are summarised in Planck Collaboration (2013b), with the general consensus being that ISW observations have constrained Ω_Λ to a value of $\Omega_\Lambda \approx 0.75 \pm 20\%$; Ω_K to be flat to within a few percent and the equation of state parameter to be $\omega \approx 1$ with no strong evolution. These results highlight the fact that the ISW effect does not constrain the ΛCDM model to anything like the precision of the standard datasets (CMB and BAOs). However, for a universe containing an amount of warm dark matter or one with a temporally varying dark energy component, the ISW effect should be an aid in constraining the models we use to describe them.

For alternative cosmological models, a variety of expectations of ISW signal arise. A study by Mainini & Mota (2012) on the effect of massive neutrinos on the ISW-LSS correlation signal, along with the expectations of different coupled dark energy models, shows that model discrimination typically involves a difference in the expected height of the peak in the cross-correlations (using the cross-correlation multiplied by $\ell(\ell+1)$ as we do here). They also note that the models are better discriminated between at higher redshifts. As redshift selection cuts modulate both the peak height and possibly the peak position of the cross-correlation signal (Figure 5.12), redshift selection effects need to be carefully deciphered.

5.4.4 Impact of using the LAV approximation

This study has focused on low-ℓ results based on the LAV approximation. As such, results for higher ℓ values ($\ell > 100$) presented here will deviate significantly from the expected ISW-induced anisotropies which include velocity information. Studies by Cai et al. (2009) and Cai et al. (2010) have looked in detail at the specific contribution the velocity information makes to the ISW anisotropies. Figure 2 in Cai et al. (2009) summarises the expected deviation of the full result from that of linear theory. Essentially, past an $\ell \sim 60$ for low redshifts ($z < 0.5$), the deviation from the full ISW anisotropies begins to become significant. This evolves to lower ℓ for higher redshifts until at $z \sim 1$ the deviation from linear theory begins to become significant at around $\ell \sim 40 - 50$. The LAV approximation, which uses full, simulated information from the density field but combines it with a linear theory velocity prescription, follows the linear theory prediction very closely at the ℓ values where the non-linear contribution becomes significant.

The effect of the non-linear component on the cross-correlation with ISW anisotropies is to suppress correlation at ℓ values that are much higher than where the non-linearities become significant in terms of raw power. Cai et al. (2009) found that the deviation from the expected linear CMB-LSS cross-correlation signal only became significant at $\ell \gtrsim 500$, which implies that all of the cross-correlation results presented here can be taken as accurate.

Chapter 6

Conclusion

This thesis has presented a number of predictions of the ΛCDM model, including the abundance of dark matter haloes across a large portion of cosmic history, observations of extreme objects in a very large simulation, and the expected ISW effect signal and its cross-correlation to LSS. In general, these predictions, generated via N-body simulations, have related to areas of the ΛCDM model that are very hard to explore. This is mainly because they all involve highly non-linear structures coupled with extreme cosmological volumes or look-back times. This places an enormous strain on the computational resources required to undertake these investigations.

6.1 Summary of results

In Chapter 3 we investigated the halo mass function and its evolution over redshift. The simulation suite analysed contained a number of high-resolution, high-redshift simulations and this allowed the mass function to be probed across an unprecedented combination of redshifts and halo masses.

The mass function from SO haloes exhibits a clear evolution with redshift, especially during the recent era of dark energy dominance ($z < 1$). We provided a redshift-parameterised fit for the SO mass function valid for the entire redshift range to within $\sim 20\%$ as well as a scheme to calculate the mass function for haloes with arbitrary overdensities. The FOF mass function displayed a weaker evolution with redshift. Because of this, we provided a universal fit for the FOF mass function, fitted to data across the entire redshift range simultaneously. We observed a modest redshift evolution in our data versus this fit.

The relative evolution of the mass functions derived via the two methods was compared.

For an SO halo defined via an overdensity of 178 versus the background matter density and an FOF halo defined via a linking length of 0.2 times the mean inter-particle separation, we find that the mass functions most closely match at $z = 0$. The disparity at $z = 0$ between the FOF and SO mass functions resides in their high mass tails where the collapsed fraction of mass in SO haloes is $\sim 80\%$ of that in FOF haloes. This difference grows with redshift so that, by $z > 20$, the SO algorithm finds a $\sim 50 - 80\%$ lower collapsed fraction in high mass haloes than the FOF algorithm.

In Chapter 4 we presented specific results from the Jubilee simulation, a large N-body, dark matter-only cosmological simulation with a volume of $V = (6\ h^{-1}\mathrm{Gpc})^3$, containing 6000^3 particles, performed within the concordance ΛCDM cosmological model. This is one of the largest simulations run to date.

The halo mass function was derived using three different halofinders and compared to fitting functions in the literature with results being consistent with previous studies across most of the mass-range of the simulation. We compared simulated clusters of maximal mass across redshifts and found that our data fit well with observed masses of extreme objects in the universe. We also explicitly confirmed that the Poisson distribution is very good at describing the distribution of extremely rare objects.

We showed the level to which cosmic variance can be expected to affect number counts of clusters in volumes smaller than $(6\ h^{-1}\mathrm{Gpc})^3$ via the halo mass function. For example, a $(500\ h^{-1}\mathrm{Mpc})^3$ volume at redshift zero would see a 10%+ error in number counts of dark matter haloes with masses equal to or greater than $4 \times 10^{14}\ h^{-1}\mathrm{M}_\odot$.

We found that objects like the Bullet cluster exist in the far-tail of the distribution of mergers in terms of relative collisional speed, perhaps implying a tension with ΛCDM. This is an intriguing result, in line with previous studies, and it adds weight to the idea that the Bullet Cluster requires further close attention.

We derived the number counts of voids in the simulation box for z =0, 0.5 and 1. Defining voids as spherical volumes containing no haloes with a mass over $10^{13}\ h^{-1}\mathrm{M}_\odot$, the void function shows a steep cut-off at a scale radius of $\sim 40\ h^{-1}\mathrm{Mpc}$. For voids defined with an aggressive mass threshold of $5 \times 10^{14}\ h^{-1}\mathrm{M}_\odot$, we find that there are no voids larger in radius than $\sim 250\ h^{-1}\mathrm{Mpc}$. These results place the lower limit on the scale at which the universe can be considered approximately homogeneous at a few hundred Mpc and they also have implications for the expected sizes of voids that we should observe in the universe in the ΛCDM model.

In Chapter 5 we presented methodologies and results on the ISW effect in the Jubilee

simulation, as well as calculating gravitational lensing outputs and mock LRGs in the same volume. As the simulation volume of the Jubilee is comparable to the Hubble volume, power on very large scales is incorporated into the ISW calculation. Haloes in the simulation are resolved down to a mass of $1.5 \times 10^{12}\ h^{-1}\mathrm{M}_{\odot}$, which allowed us to derive a catalogue of mock Luminous Red Galaxies (LRGs) for cross-correlation analysis with the ISW signal.

We found the ISW effect to be strongest in a redshift range of $z \sim 0.2 - 0.8$, with the evolution of the ISW power spectrum matching expectations from linear theory. Maps of the gravitational lensing effect, including the convergence and deflection fields, were calculated using the same potential as for the ISW. We calculated the redshift dependence of the ISW-LRG cross-correlation signal. The signal was found to be strongest in LSS surveys that can probe the redshift range $z = 0.2$ to 0.8, matching the ISW effect peak. There is evidence that the signal peaks at larger angular scales ($\ell \sim 10$) for lower redshifts and moves to smaller angular scales for higher redshifts.

6.2 Avenues for future work

There are many areas of future study on which this thesis touches.

The methodology and results of Chapter 3 highlight the current state of studies into the abundance of dark matter haloes in the universe. This study concluded with some comments on what future research needs to be undertaken in order for the mass function to be constrained to a greater accuracy. They can be briefly summarised as: understanding the interplay of baryons and dark matter on the mass function; understanding the influence of halo definitions on the mass function and its subsequent evolution, including how the definitions lead to pseudo-mass growth; understanding the effect of mergers on the mass function. There is also the need to run larger and larger simulations in order to study the mass function with precision. An alternative, or perhaps complementary, approach is to run multiple realisations of boxes and also boxes with different specifications that interleave with each other in terms of mass and volume coverage. Because of the reliance on simulations in this particular topic, it is likely that papers will periodically be issued for some time to come, each one honing in, it is hoped, on a more precise function that observers and theorists can then utilise in their studies.

The question of the velocity of the Bullet Cluster is still an open one and clearly there are more careful follow-up studies that can be performed. The methodology we used to investigate the Bullet Cluster candidates in the Jubilee volume can certainly be improved.

In addition, a logical avenue of study which should be undertaken is the extension of the Extreme Value Statistics approach to the distribution of the velocities of objects. In the light of the Extreme Value Statistics framework the Bullet Cluster may not appear so challenging to ΛCDM.

The Jubilee simulation has a large enough volume that statistics on local measurements of the Hubble parameter can be investigated. Taking observers at either the positions of haloes or at random locations in the simulation volume, we can ask the question: what recession velocities do they observe for nearby galaxies? For distances up to a few hundred Mpc we can build up statistics in independent volumes and then relate the findings to what we observe in reality. A study along these lines would be useful in attempting to understand systematic errors in local measurements of the Hubble parameter.

Much of the work presented in Chapter 5 represents the building of important datasets from which further ISW and weak-lensing science can be undertaken. Work is continuing at the time of writing on a number of the different areas touched upon in that chapter. Perhaps the most important avenue of research based on the ISW results is that of moving from signal-space for the ISW-LSS cross-correlation into detection space. To attempt this, a mock CMB map, including the ISW effect from the Jubilee simulation, needs to be constructed. Further to this, particulars of LSS surveys need to be overlaid on the mock LRGs, including considerations such as sky coverage and catalogue completeness. Also being undertaken is the creation of a mock catalogue of radio galaxies to match that of the NVSS, which has been employed extensively in ISW detections to date.

The lensing maps shown in Chapter 5 are, in the context of this work, a bi-product of the gravitational potential that has been calculated in the simulation for studying the ISW. Despite the lack of emphasis in this thesis on this data product there is a significant amount of follow-up work on it that can now be undertaken. Initially, there is the application of the lensing information to the question of the ISW-lensing cross-correlation, in a similar manner to that applied by the Planck collaboration (Planck Collaboration, 2013b), which was briefly discussed in Chapter 5. There is also the possibility of exploiting the lensing maps to investigate cosmic shear with a view to optimising future LSS surveys.

Finally an important question to answer, which will be facilitated by the Jubilee data, is that of the anomalously strong stacking-derived ISW results discussed briefly in Chapter 5. Using the Jubilee simulation we plan on thoroughly analysing the various subtleties that arise in the stacking approach and judge just how irregular the result is. We also hope that an optimal method for detecting a ΛCDM ISW effect via stacking can be constructed

from the data.

We are led to believe a lie
When we see not thro' the eye,
Which was born in a night to perish in a night,
When the soul slept in beams of light.

– William Blake

Bibliography

Aarseth S. J., 1963, MNRAS, 126, 223 34

Afshordi N., Loh Y. S., Strauss M. A., 2004, Phys Rev D, 69, 083524 102, 103

Ahn C. P. et al., 2012, ApJs, 203, 21 131

Alexander S., Biswas T., Notari A., Vaid D., 2009, JCAP, 9, 025 100

Amiaux J. et al., 2012. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 8442 5, 8, 82

Anderson L. et al., 2012, MNRAS, 427, 3435 4

Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012, MNRAS, 426, 2046 46, 49, 70, 72, 88, 89

Athanassoula E., Fady E., Lambert J. C., Bosma A., 2000, MNRAS, 314, 475 36

Audit E., Teyssier R., Alimi J. M., 1998, A.&A, 333, 779 76

Bagla J., Khandai N., Kulkarni G., 2009, arXiv:0908.2702 55

Bagla J. S., Prasad J., 2006, MNRAS, 370, 993 53, 55, 91

Balaguera-Antolínez A., Sánchez A. G., Böhringer H., Collins C., Guzzo L., Phleps S., 2011, MNRAS, 413, 386 45

Barausse E., Matarrese S., Riotto A., 2005, Phys Rev D, 71, 063537 100

Barber A. J., Thomas P. A., Couchman H. M. P., 1999, MNRAS, 310, 453 104

Barkana R., Loeb A., 2004, ApJ, 609, 474 57, 89

Barrena R., Biviano A., Ramella M., Falco E. E., Seitz S., 2002, A.&A, 386, 816 93

Behroozi P. S., Wechsler R. H., Wu H. Y., 2013, ApJ, 762, 109 112

Benítez N. et al., 2009, ApJ, 691, 241 5, 82

Bergström L., 2012, Annalen der Physik, 524, 479 5, 26

Bertschinger E., 1985, ApJs, 58, 39 43

Bertschinger E., 1995, ArXiv Astrophysics e-prints 37

Bertschinger E., 1998, *Ann. Rev. Astron.& Astrophys.* , 36, 599 34

Bhattacharya S., Heitmann K., White M., Lukic Z., Wagner C. et al., 2011, ApJ, 732, 122 46, 47, 53, 57, 70, 72, 75

Böhringer H. et al., 2004, A.&A, 425, 367 4

Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440 46, 54

Boughn S., Crittenden R., 2004, Nature, 427, 45 103

Boughn S. P., Crittenden R. G., 2002, Physical Review Letters, 88, 021302 103

Bradač M. et al., 2006, ApJ, 652, 937 93

Buchert T., 1993, A.&A, 267, L51 38

Burenin R. A., Vikhlinin A., Hornstrup A., Ebeling H., Quintana H., Mescheryakov A., 2007, ApJs, 172, 561 4

Cabré A., Fosalba P., Gaztañaga E., Manera M., 2007, MNRAS, 381, 1347 131

Cai Y. C., Cole S., Jenkins A., Frenk C., 2009, MNRAS, 396, 772 135

Cai Y. C., Cole S., Jenkins A., Frenk C. S., 2010, MNRAS, 407, 201 104, 105, 122, 133, 135

Caldwell R. R., Kamionkowski M., Weinberg N. N., 2003, Physical Review Letters, 91, 071301 21

Carbone C., Springel V., Baccigalupi C., Bartelmann M., Matarrese S., 2008, MNRAS, 388, 1618 104

Carbone C., Baldi M., Pettorino V., Baccigalupi C., 2013, ArXiv e-prints 104

Cayón L., Gordon C., Silk J., 2011, MNRAS, 415, 849 83

Chisari N. E., Zaldarriaga M., 2011, Phys Rev D, 83, 123505 27

Clowe D., Gonzalez A., Markevitch M., 2004, ApJ, 604, 596 93

Clowe D., Bradač M., Gonzalez A. H., Markevitch M., Randall S. W., Jones C., Zaritsky D., 2006, ApJL, 648, L109 93

Cohn J. D., White M., 2008, MNRAS, 385, 2025 46, 78

Coil A. L. et al., 2011, ApJ, 741, 8 45

Colberg J. M. et al., 2008, MNRAS, 387, 933 84

Cole S., Lacey C., 1996, MNRAS, 281, 716 47, 81

Condon J. J., Cotton W. D., Greisen E. W., Yin Q. F., Perley R. A., Taylor G. B., Broderick J. J., 1998, AJ, 115, 1693 3, 131

Conley A. et al., 2011, ApJs, 192, 1 4

Cooray A., 2002, Phys Rev D, 65, 103510 105

Cooray A., Sheth R. K., 2002, Phys. Rept., 372, 1 6

Corasaniti P., Achitouv I., 2011, Phys. Rev., D84, 023009 46

Courtin J., Rasera Y., Alimi J. M., Corasaniti P. S., Boucher V. et al., 2011, MNRAS, 410, 1911 46, 47, 48, 49, 70, 72, 75, 78, 81

Crittenden R. G., Turok N., 1996, Physical Review Letters, 76, 575 102

Crocce M., Pueblas S., Scoccimarro R., 2006, MNRAS, 373, 369 38

Crocce M., Fosalba P., Castander F. J., Gaztañaga E., 2010, MNRAS, 403, 1353 83

Crocce M., Fosalba P., Castander F. J., Gaztañaga E., 2010, MNRAS, 403, 1353 38, 46, 47, 49, 70, 72, 75

Cui W., Borgani S., Dolag K., Murante G., Tornatore L., 2012, MNRAS, 423, 2279 50

Das S., Bode P., 2008, ApJ, 682, 1 104

Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371 43, 46, 47, 81

Dawson K. S. et al., 2013, AJ, 145, 10 5, 45, 82

Dehnen W., 2001, MNRAS, 324, 273 36

Dehnen W., Read J. I., 2011, European Physical Journal Plus, 126, 55 34, 35, 36

Diemer B., More S., Kravtsov A., 2012, arXiv:1207.0816 49

Douspis M., Castro P. G., Caprini C., Aghanim N., 2008, A.&A, 485, 395 103

Drinkwater M. J. et al., 2010, MNRAS, 401, 1429 5, 82

Driver S. P. et al., 2011, MNRAS, 413, 971 45

Dunkley J. et al., 2009, ApJs, 180, 306 41

Dyer C. C., Ip P. S. S., 1993, ApJ, 409, 60 36

Eales S. et al., 2010, PASP, 122, 499 4

Efstathiou G., 1979, Ph.D. thesis, Durham 36

Einasto J., Einasto M., Gramann M., 1989, MNRAS, 238, 155 85

Eisenstein D. J. et al., 2005, ApJ, 633, 560 108, 110, 112, 115

Eke V. R., Cole S., Frenk C. S., 1996, MNRAS, 282, 263 49

Eke V. R., Baugh C., Cole S., Frenk C., Navarro J., 2006, MNRAS, 370, 1147 6

Ellis G. F. R., 1979, General Relativity and Gravitation, 11, 281 100

Fakhouri O., Ma C. P., 2010, MNRAS, 401, 2245 49

Fan X., 2012, Research in Astronomy and Astrophysics, 12, 865 6

Farrar G. R., Rosen R. A., 2007, Physical Review Letters, 98, 171302 94

February S., Larena J., Smith M., Clarkson C., 2010, MNRAS, 405, 2231 100

Flender S., Hotchkiss S., Nadathur S., 2013, JCAP, 2, 013 103

Forero-Romero J. E., Gottlöber S., Yepes G., 2010, ApJ, 725, 598 94

Fosalba P., Gaztañaga E., Castander F. J., 2003, ApJL, 597, L89 103

Fosalba P., Gaztañaga E., Castander F. J., Manera M., 2008, MNRAS, 391, 435 104

Fowler J. W. et al., 2010, ApJ, 722, 1148 45

Freedman W. L. et al., 2001, ApJ, 553, 47 19

Gao L., Navarro J. F., Cole S., Frenk C. S., White S. D. M., Springel V., Jenkins A., Neto A. F., 2008, MNRAS, 387, 536 78

Gaztañaga E., Manera M., Multamäki T., 2006, MNRAS, 365, 171 134

Giannantonio T., Scranton R., Crittenden R. G., Nichol R. C., Boughn S. P., Myers A. D., Richards G. T., 2008, Phys Rev D, 77, 123520 103, 134

Gilbank D. G., Gladders M. D., Yee H. K. C., Hsieh B. C., 2011, AJ, 141, 94 45

Gill S. P., Knebe A., Gibson B. K., 2004, MNRAS, 351, 399 42

Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759 v, 107

Gottlöber S., Łokas E. L., Klypin A., Hoffman Y., 2003, MNRAS, 344, 715 85

Granett B. R., Neyrinck M. C., Szapudi I., 2008, ApJL, 683, L99 103, 132, 133

Green S. R., Wald R. M., 2012, Phys Rev D, 85, 063512 27

Gunn J. E., Gott III J. R., 1972, ApJ, 176, 1 29

Guth A. H., 1981, Phys Rev D, 23, 347 23

Guy J. et al., 2007, A.&A, 466, 11 4

Guy J. et al., 2010, A.&A, 523, A7 4

Haggerty M. J., Janin G., 1974, A.&A, 36, 415 34

Haiman Z., Mohr J. J., Holder G. P., 2001, arXiv:astro-ph/0103049 6

Harnois-Deraps J., Pen U. L., Iliev I. T., Merz H., Emberson J. et al., 2012, arXiv:1208.5098 37, 39, 40, 42, 43

Harrison I., Coles P., 2011, MNRAS, 418, L20 83, 91, 92

Harrison I., Coles P., 2012, MNRAS, 421, L19 6

Harrison I., Coles P., 2012, MNRAS, 421, L19 83, 91, 92

Harrison I., Hotchkiss S., 2012, arXiv:1210.4369 91

Hasselfield M. et al., 2013, arXiv:1301.0816 45

Haugg T., Hofmann S., Kopp M., 2012, arXiv:1211.0011 27

Hayashi E., White S. D. M., 2006, MNRAS, 370, L38 94, 95

Healey S. E., Romani R. W., Taylor G. B., Sadler E. M., Ricci R., Murphy T., Ulvestad J. S., Winn J. N., 2007, ApJs, 171, 61 4

Heath D. J., 1977, MNRAS, 179, 351 50, 106

Heinrich J., 2003, CDF, Note 6438 52

Heitmann K., Ricker P. M., Warren M. S., Habib S., 2005, ApJs, 160, 28 38

Heitmann K., Lukić Z., Habib S., Ricker P. M., 2006, ApJL, 642, L85 46, 53

Heitmann K. et al., 2008, Computational Science and Discovery, 1, 015003 38

Hernandez-Monteagudo C., Smith R. E., 2012, arXiv:1212.1174 104, 132, 133

Heymans C. et al., 2012, MNRAS, 427, 146 4

Hilbert S., Hartlap J., White S. D. M., Schneider P., 2009, A.&A, 499, 31 104

Hinshaw G. et al., 2012, arXiv:1212.5226 3

Hobson M. P., Efstathiou G., Lasenby A. N., 2006, General Relativity: An Introduction for Physicists. Cambridge University Press 11

Hockney R. W., Eastwood J. W., 1988, Computer Simulation Using Particles. Adam Hilger Ltd., Bristol, UK 36, 107

Hoekstra H., Jain B., 2008, Annual Review of Nuclear and Particle Science, 58, 99 104

Hogg D. W., 1999, arXiv:astro-ph/9905116 18, 19

Holder G., Haiman Z., Mohr J., 2001, ApJ, 560, L111 6

Holmberg E., 1941, ApJ, 94, 385 34

Holtzman J. A. et al., 2008, AJ, 136, 2306 4

Holz D. E., Perlmutter S., 2012, ApJL, 755, L36 83

Hoyle B., Jimenez R., Verde L., 2011, Phys Rev D, 83, 103502 83

Hu W., Sugiyama N., 1994, Phys Rev D, 50, 627 102

Hubble E., 1929a, Proceedings of the National Academy of Science, 15, 168 2

Hubble E. P., 1929b, ApJ, 69, 103 2

Iguchi H., Nakamura T., Nakao K., 2002, Progress of Theoretical Physics, 108, 809 100

Ilic S., Langer M., Douspis M., 2013, arXiv:1301.5849 104, 132, 133

Iliev I. T., Shapiro P. R., 2001, MNRAS, 325, 468 43, 76

Iliev I. T., Shapiro P. R., Mellema G., Merz H., Pen U. L., 2008, in refereed proceedings of TeraGrid08, ArXiv e-prints (0806.2887) 39

Iliev I. T., Ahn K., Koda J., Shapiro P. R., Pen U. L., 2010, proceedings paper for Moriond 2010 meeting (ArXiv:1005.2502) 39

Jain B., Seljak U., White S., 2000, ApJ, 530, 547 104

Jeans J. H., 1902, Royal Society of London Philosophical Transactions Series A, 199, 1 28

Jeans J. H., 1915, MNRAS, 76, 70 35

Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, MNRAS, 321, 372 46, 47, 48, 49, 50, 70, 75

Jimenez R., Verde L., 2009, Phys Rev D, 80, 127302 83

Jing Y. P., 2005, ApJ, 620, 559 85

Kamionkowski M., Spergel D. N., 1994, ApJ, 432, 7 103

Karachentsev I. D., Karachentseva V. E., Huchtmeier W. K., Makarov D. I., 2004, AJ, 127, 2031 2

Kazin E. A. et al., 2010, ApJ, 710, 1444 110, 111

Kessler R. et al., 2009, ApJs, 185, 32 4

Kiessling A., Heavens A. F., Taylor A. N., Joachimi B., 2011, MNRAS, 414, 2235 104

Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ, 740, 102 46

Knebe A., Wagner C., Knollmann S., Diekershoff T., Krause F., 2009, ApJ, 698, 266 38

Knebe A., Knollmann S. R., Muldrew S. I., Pearce F. R., Aragon-Calvo M. A. et al., 2011, arXiv:1104.0949 42, 96

Knollmann S. R., Knebe A., 2009, ApJs, 182, 608 42, 43

Komatsu E. et al., 2009, ApJs, 180, 330 41, 51

Kravtsov A., Borgani S., 2012, arXiv:1205.5556 6

Kuhlen M., Vogelsberger M., Angulo R., 2012, Physics of the Dark Universe, 1, 50 34, 37, 83

Lacey C., Cole S., 1993, MNRAS, 262, 627 41, 49

Lacey C., Cole S., 1994, MNRAS, 271, 676 46, 76

Lahav O., Lilje P. B., Primack J. R., Rees M. J., 1991, MNRAS, 251, 128 49

Lawrence E., Heitmann K., White M., Higdon D., Wagner C., Habib S., Williams B., 2010, ApJ, 713, 1322 104

Lee J., Komatsu E., 2010, ApJ, 718, 60 83, 94, 95

Lee J., Shandarin S. F., 1998, arXiv:astro-ph/9811004 46

Lemaître G., 1927, Annales de la Societe scientifique de Bruxelles, 47, 49 2

Lewis A., Challinor A., Lasenby A., 2000, Astrophys. J., 538, 473 37, 41

Lim S., Lee J., 2012, arXiv:1206.5351 46

Llinares C., Zhao H. S., Knebe A., 2009, ApJL, 695, L145 94

LSST Dark Energy Science Collaboration, 2012, arXiv:1211.0310 5, 45, 82

Lukic Z., Heitmann K., Habib S., Bashinsky S., Ricker P. M., 2007, ApJ, 671, 1160 38, 39, 46, 52, 53, 55, 57, 58, 75, 91

Lukic Z., Reed D., Habib S., Heitmann K., 2009, ApJ, 692, 217 47, 57, 78, 81

Ma C., Bertschinger E., 1995, ApJ, 455, 7 47, 81

Maggiore M., Riotto A., 2010, ApJ, 711, 907 46

Magnier E. A. et al., 2013, ApJs, 205, 20 5, 82

Mainini R., Mota D. F., 2012, ApJ, 744, 3 134

Mandelbaum R., Slosar A., Baldauf T., Seljak U., Hirata C. M., Nakajima R., Reyes R., Smith R. E., 2013, MNRAS, in print, arXiv:1207.1120 8, 104

Markevitch M., Gonzalez A. H., David L., Vikhlinin A., Murray S., Forman W., Jones C., Tucker W., 2002, ApJL, 567, L27 93

Markevitch M., Gonzalez A. H., Clowe D., Vikhlinin A., Forman W., Jones C., Murray S., Tucker W., 2004, ApJ, 606, 819 95, 97

Marra V., Pääkkönen M., 2010, JCAP, 12, 021 100

Martel H., 2005, arXiv:astro-ph/0506540 37

Mastropietro C., Burkert A., 2008, MNRAS, 389, 967 94

McEwen J. D., Vielva P., Hobson M. P., Martínez-González E., Lasenby A. N., 2007, MNRAS, 376, 1211 103

Mehrtens N. et al., 2012, MNRAS, 423, 1024 45

Milosavljević M., Koda J., Nagai D., Nakar E., Shapiro P. R., 2007, ApJL, 661, L131 94

Moffat J. W., 2005, JCAP, 10, 012 100

Mohr J. J. et al., 2012, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. Vol. 8451 5, 82

More S., 2011, ApJ, 741, 19 45

More S., Kravtsov A., Dalal N., Gottlober S., 2011, ApJs, 195, 4 47, 48, 57, 58, 76, 78, 81

Moutarde F., Alimi J. M., Bouchet F. R., Pellat R., Ramani A., 1991, ApJ, 382, 377 38

Müller V., Arbabi-Bidgoli S., Einasto J., Tucker D., 2000, MNRAS, 318, 280 101

Mullis C. R., Rosati P., Lamer G., Böhringer H., Schwope A., Schuecker P., Fassbender R., 2005, ApJL, 623, L85 83

Murphy T. et al., 2010, MNRAS, 402, 2403 4

Mustapha N., Hellaby C., Ellis G. F. R., 1997, MNRAS, 292, 817 100

Nadathur S., Sarkar S., 2011, Phys Rev D, 83, 063506 100

Nadathur S., Hotchkiss S., Sarkar S., 2012, JCAP, 6, 042 103, 132, 133

Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493 43, 76

Navarro J. F. et al., 2004, MNRAS, 349, 1039 78

Navarro J. F. et al., 2010, MNRAS, 402, 21 78

Nolta M. R. et al., 2004, ApJ, 608, 10 103

Norris R. P., 2011, Journal of Astrophysics and Astronomy, 32, 599 132

Nusser A., 2008, MNRAS, 384, 343 94

Opik E., 1922, ApJ, 55, 406 1

Padmanabhan N., Hirata C. M., Seljak U., Schlegel D. J., Brinkmann J., Schneider D. P., 2005, Phys Rev D, 72, 043525 133

Padmanabhan T., 1993, Structure formation in the universe. Cambridge University Press 27, 28, 30, 32

Penzias A. A., Wilson R. W., 1965, ApJ, 142, 419 3

Perlmutter S. et al., 1999, ApJ, 517, 565 3, 19

Pietrobon D., Balbi A., Marinucci D., 2006, Phys Rev D, 74, 043524 134

Pillepich A., Porciani C., Reiprich T. H., 2012, MNRAS, 422, 44 45

Planck Collaboration, 2011a, A.&A, 536, A8 45

Planck Collaboration, 2011b, A.&A, 536, A10 46

Planck Collaboration, 2011c, A.&A, 536, A11 46

Planck Collaboration, 2013a, arXiv:1303.5062 3, 7, 102

Planck Collaboration, 2013b, arXiv:1303.5079 7, 103, 104, 131, 133, 134, 139

Planck Collaboration, 2013c, arXiv:1303.5076 20, 21, 26, 32, 41, 103

Planck Collaboration, 2013d, arXiv:1303.5089 45, 46

Power C., Knebe A., 2006, MNRAS, 370, 691 53, 89

Prada F., Klypin A. A., Cuesta A. J., Betancort-Rijo J. E., Primack J., 2012, MNRAS, 423, 3018 81, 89

Press W. H., Schechter P., 1974, ApJ, 187, 425 34, 46, 49, 70, 72

Rawlings S., 2011, arXiv:1105.6333 6

Reed D., Gardner J., Quinn T. R., Stadel J., Fardal M. et al., 2003, MNRAS, 346, 565 49, 70, 72

Reed D., Bower R., Frenk C., Jenkins A., Theuns T., 2007, MNRAS, 374, 2 46, 47, 49, 55, 65, 70, 75

Reed D. S., Smith R. E., Potter D., Schneider A., Stadel J. et al., 2012, arXiv:1206.5302 38, 58

Rees M. J., Sciama D. W., 1968, Nature, 217, 511 102

Riess A. G. et al., 1998, AJ, 116, 1009 3, 19

Rines K., Diaferio A., Natarajan P., 2008, ApJL, 679, L1 6

Rosati P. et al., 2009, A.&A, 508, 583 83

Röttgering H. et al., 2011, Journal of Astrophysics and Astronomy, 32, 557 132

Rozo E., Bartlett J. G., Evrard A. E., Rykoff E. S., 2012a, arXiv:1204.6305 46

Rozo E., Vikhlinin A., More S., 2012b, ApJ, 760, 67 46

Rozo E. et al., 2010, ApJ, 708, 645 6, 45

Sachs R. K., Wolfe A. M., 1967, ApJ, 147, 73 102, 103, 105

Sarkar P., Yadav J., Pandey B., Bharadwaj S., 2009, MNRAS, 399, L128 20, 98

Schiavon F., Finelli F., Gruppuso A., Marcos-Caballero A., Vielva P., Crittenden R. G., Barreiro R. B., Martínez-González E., 2012, MNRAS, 427, 3044 103

Schlegel D. J. et al., 2009, arXiv:0904.0468 5, 82

Seljak U., Zaldarriaga M., 1996, ApJ, 469, 437 37

Shapiro P. R., Iliev I. T., Raga A. C., 1999, MNRAS, 307, 203 43, 76

Sheth R. K., Tormen G., 2002, MNRAS, 329, 61 46, 49, 70, 72, 77

Sirko E., 2005, ApJ, 634, 728 53, 89

Smith R. E., Marian L., 2011, arXiv:1106.1665 6, 89, 93

Smoot G. F. et al., 1992, ApJL, 396, L1 3

Springel V., 2005, MNRAS, 364, 1105 38, 42, 43

Springel V., Farrar G. R., 2007, MNRAS, 380, 911 94, 95, 97

Springel V. et al., 2005, Nature, 435, 629 89

Stadel J. G., 2001, Ph.D. thesis, University of Washington 38

Summers F. J., Davis M., Evrard A. E., 1995, ApJ, 454, 1 76

Suzuki N. et al., 2012, ApJ, 746, 85 4

Sylos Labini F., Baryshev Y. V., 2010, JCAP, 6, 021 20, 98

Takey A., Schwope A., Lamer G., 2011, A.&A, 534, A120 45

Takizawa M., 2005, ApJ, 629, 791 94

Takizawa M., 2006, PASJ, 58, 925 94

Teyssier R. et al., 2009, A.&A, 497, 335 89, 104

The COrE Collaboration et al., 2011, ArXiv e-prints 8

Thompson R., Nagamine K., 2012, MNRAS, 419, 3560 94, 95

Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, ApJ, 688, 709 55, 87, 88

Tinker J. L., Kravtsov A. V., Klypin A., Abazajian K., Warren M. S. et al., 2008, ApJ, 688, 709 x, 38, 46, 47, 58, 62, 64, 65, 72, 74, 75, 76, 91

Tomita K., 2001, MNRAS, 326, 287 100

Vale C., White M., 2003, ApJ, 592, 699 104

Varela J. et al., 2009, A.&A, 497, 667 45

Vielva P., Martínez-González E., Tucci M., 2006, MNRAS, 365, 891 103

Vikhlinin A. et al., 2009, ApJ, 692, 1060 6, 45, 46

von Hoerner S., 1960, Z. Astrophys., 50, 184 34

Waizmann J. C., Redlich M., Bartelmann M., 2012, A.&A, 547, A67 83

Warren M. S., Abazajian K., Holz D. E., Teodoro L., 2006, ApJ, 646, 881 53, 57, 58, 70, 75, 76

Watson W. A., Iliev I. T., D'Aloisio A., Knebe A., Shapiro P. R., Yepes G., 2012, arXiv:1212.0095 45

Watson W. A., Iliev I. T., Diego J. M., Gottlöber S., Knebe A., Martínez-González E., Yepes G., 2013a, arXiv:1305.1976 82

Watson W. A. et al., 2013b, arXiv:1307.1712 102

Weller J., Battye R., Kneissl R., 2002, Phys. Rev. Lett., 88, 231301 6

Wen Z. L., Han J. L., Liu F. S., 2012, ApJs, 199, 34 108

White M., 2001, A.&A, 367, 27 47, 81

White M., 2002, ApJs, 143, 241 47, 75, 81

Wiltshire D. L., 2005, arXiv:gr-qc/0503099 100

Yadav J., Bharadwaj S., Pandey B., Seshadri T. R., 2005, MNRAS, 364, 601 20, 98

Yoshida N., Sokasian A., Hernquist L., Springel V., 2003, ApJ, 598, 73 55, 89

Zehavi I., Riess A. G., Kirshner R. P., Dekel A., 1998, ApJ, 503, 483 100

Zel'dovich Y. B., 1970, A.&A, 5, 84 38, 39

Zheng Z., Zehavi I., Eisenstein D. J., Weinberg D. H., Jing Y. P., 2009, ApJ, 707, 554 107, 108, 109, 110, 112

Zheng Z. et al., 2005, ApJ, 633, 791 108

Zitrin A., Bartelmann M., Umetsu K., Oguri M., Broadhurst T., 2012, MNRAS, 426, 2944 108, 110, 112, 113

Zwicky F., 1933, Helvetica Physica Acta, 6, 110 2